

Evaluating Motivational Interviewing Measures of Knowledge and Skill Using Training
Outcome Data

A thesis submitted in partial fulfilment of the requirements for the

Degree of

Master of Science in Psychology

University of Canterbury

Emma Jane Patricia Dear

2014

Supervisors:

Dr Eileen Britt, Department of Psychology

Professor Randolph Grace, Department of Psychology

Table of Contents

List of Tables.....	4
List of Figures	4
Abbreviations	5
MISC and MITI summary score abbreviations	5
VASE-R subcomponent abbreviations	6
Acknowledgements	7
Abstract	8
Introduction.....	9
A brief history of Motivational Interviewing	11
What is Motivational Interviewing?	11
Motivational Interviewing Efficacy.....	13
Further Development of Motivational Interviewing	15
Implementation Training	26
The Training of Motivational Interviewing.....	28
Evaluating Motivational Interviewing Training	31
Method	69
Procedure	69
Participants	71
Implementation Plan.....	72
Measures	73

Data Analysis.....	75
Results	83
The MIKAT	83
The VASE-R.....	89
Discussion	95
Summary of Findings	96
Recommendations	101
Limitations and Future Directions	109
References	112
Appendix B	122
Appendix C	123
Appendix D	141
Appendix E.....	145

List of Tables

Table 1. Participant demographic characteristics	72
Table 2. Means and Standard Deviations for Items, Components, and Total MIKAT scores for Pre- and Post-Training	84
Table 3. Internal Consistency of the Total, Component, and Subcomponent MIKAT Scores for Pre- and Post-Training	85
Table 4. Scale correlations of the MIKAT Components and Subcomponents for Pre- and Post-Training.....	86
Table 5. Component Loadings of the MIKAT.....	87
Table 6. Means and Standard Deviations for VASE-R Items, Subscales, and Total Scores for Pre- and Post-Training	90
Table 7. Internal Consistency of the Total and Subscale VASE-R Scores for Pre- and Post- Training.....	91
Table 8. Disattenuated scale correlations of the VASE-R Subscales for Pre- and Post- Training.....	92
Table 9. Eigenvalues Generated from the Parallel Analysis: the Actual Eigenvalues Correspond to the VASE-R whereas the Mean and 95 th Percentile were Generated at Random.....	93

List of Figures

<i>Figure 1.</i> Discrepancy histogram displaying the similarities in VASE-R point changes from pre- to post-training. VASE-R = Video Assessment of Simulated Encounters–Revised.....	94
<i>Figure 2.</i> An example of the recommended paired-items format to replace the true–false component.....	103
<i>Figure 3.</i> A further example of the recommended paired-items format, this time using items from the original MIKAT.	104

Abbreviations

MI: Motivational Interviewing

MINT: Motivational Interviewing Network of Trainers

MISC: The Motivational Interviewing Skills Code

MITI: The Motivational Interviewing Treatment Integrity (Code)

MIKAT: The Motivational Interviewing Knowledge and Attitudes Test

HRQ: The Helpful Responses Questionnaire

ORQ: Officer Response Questionnaire

WRQ: Worker's Response Questionnaire

VASE-R: The Video Assessment of Simulated Encounters – Revised

ICCs: Intra-class correlations

CYFs: Child Youth and Family services

MCAR: Little's missing completely at Random Test

MISC and MITI summary score abbreviations

%OQ: Percentage of open-ended questions

%CR: Percentage of complex reflections

R: Q: Reflection to question ratio

MI-Con: MI consistent responses (MISC)

MI-In: MI inconsistent responses (MISC)

%CCT: Percentage of client change talk (MISC)

MI-Ad: MI-adherent behaviour (MITI)

VASE-R subcomponent abbreviations

RL: Reflective listening

RR: Responding to resistance

ECT: Eliciting change talk

S: Summary

DD: Developing discrepancy

Acknowledgements

I would like to thank my supervisors Eileen Britt and Randolph Grace for their guidance and expertise. Their commitment and patience during this occasionally “rocky” experience was great.

Big thanks to Robert, the statistical advisor for his guidance through some rather tough analyses. His help with the many statistical analyses required for the present paper was invaluable and very greatly appreciated.

My appreciation is further extended to my parents (Richard and Helen) and my partner (Slava) for their continued support throughout my many years at University. I do not know what I would have done without them.

Abstract

The efficacy of Motivational Interviewing (MI), like other evidence-based counselling interventions, relies on the skills of the practitioner. Therefore, it is imperative that such practitioners receive sufficient training in order to successfully implement the method. Hence, it is important to ensure a given training programme has been effective through the use of reliable and valid measures of knowledge and skill acquisition. The Motivational Interviewing Treatment Integrity code (MITI) is one such measure of MI skill attainment; however its use is labour intensive and requires the submission of in-session trainee audio recordings. Other measures of MI skill or knowledge also exist which may be more practical for evaluating training programmes. However, these measures are under researched and are therefore rarely utilised across the MI training literature. The present research is an investigation of the psychometric properties of the Motivational Interviewing Knowledge and Attitudes Test (MIKAT), a test of MI knowledge and attitudes, and the Video Assessment of Simulated Encounters–Revised (VASE-R), a test of MI skill attainment through the use of video to simulate encounters with clients. The measures were administered during a training programme for staff of Child Youth and Family services (CYFs) New Zealand. The results suggest that the MIKAT and VASE-R are sensitive to the effects of training. However, it appeared both measures require refinement and possibly restructuring to make them more consistent with the most recent MI developments.

Evaluating Under Researched Motivational Interviewing Measures of Knowledge and Skill Using Training Outcome Data

Motivational Interviewing (MI) was developed by Miller and Rollnick (1991; 2002; 2012) as a counselling method that utilizes the exploration and enhancement of intrinsic motivation as a way of facilitating change. The method has been demonstrated across diverse client populations to be effective in promoting behaviour change (e.g., Lundahl, Kunz, Brownell, Tollefson & Burke, 2010). It is clear that, like other evidence-based methods, the efficacy of MI is reliant on the skills of the practitioner, therefore sufficient training is imperative. In order to determine whether a given training programme has been effective, evaluations of trainee skill attainment need to be evaluated using reliable and valid measures.

Presently there are only a handful of established measures of MI knowledge, skill, adherence and performance. Amongst these only two have received adequate psychometric investigations and have been deemed reliable and valid. Unfortunately these measures are also the most labour-intensive, requiring the evaluator to listen to in-session recordings and rate a practitioner on their use of MI. Less time-consuming written-based measures of MI knowledge and skill could be useful for efficiently evaluating MI training without the need to listen to individual in-session recordings. However, the other measures of MI skill or knowledge have been subjected to little psychometric evaluation and have therefore been rarely used across the literature. Full psychometric evaluations would provide support for the reliability and validity of these measures and therefore determine whether they would be useful resources for future training evaluations.

The present study evaluated two under researched MI measures in order to determine whether they were in fact resources that could be used to make training evaluations more efficient. The chosen measures were the Motivational Interviewing Knowledge and Attitudes Test (MIKAT), which assess MI knowledge and attitudes and the Video Assessment of

Simulated Encounters–Revised (VASE-R), a test of MI skill. Data was provided by a training programme for staff of Child Youth and Family services (CYFs) New Zealand with permission to carry out full psychometric evaluations of the measures. The major question was whether the MIKAT and VASE-R were reliable and valid measures of MI knowledge and skill. If the measures are in fact reliable and valid, evaluators of future studies may consider utilising them to efficiently provide a record of training progression. The measures could also be used alongside the more labour-intensive tests in order to develop a more comprehensive evaluation of a given training.

The first sections of the introduction provide a detailed background of MI itself. After acknowledging the origin of MI, an overview of the method along with details of efficacy outcome research is provided. Recently MI has undergone significant developments that included revisions to both the structure and spirit (the underlying perspective with which one practices MI, Miller & Rollnick, 2012); therefore these developments and thorough accounts of the current MI structure and spirit are included.

Following the background of MI the focus on training becomes apparent. What constitutes an effective training and implementation plan is followed by Miller and Rollnick's (2012) guidelines to MI training. All currently established measures of MI knowledge, skill, adherence, and performance will then be reviewed. Thorough descriptions along with accounts of psychometric investigations and summaries of outcome research utilising each measure are provided. From this section the reader will be able to appreciate the rare use of other measures of MI skill or knowledge and therefore the need for investigations and further development to increase their use within the literature and perhaps make evaluations of MI training more efficient. Finally, a brief background and description of the CYFs training programme is provided and the rationale for the present research is given.

A brief history of Motivational Interviewing

The origin of MI goes back to when Miller was alerted to the significant impact of practitioner empathy on behaviour change. Practitioners were observed in-session with problem drinking outpatients and were rated on their empathetic understanding (Miller, Taylor, & West, 1980). This empathetic understanding predicted 67% of the variance in practitioner success rate (decreased problem drinking). This relationship, despite a decline in strength, was sustained over a two-year follow-up: Empathetic understanding predicted half of the variance at 12 months, and one quarter at 24 months (Miller & Baca, 1983). These results, paired with the knowledge that practitioner behaviours associated with harsh confrontation were likely to result in adverse outcomes for the client (Lieberman, Yalom & Miles, 1973, as cited in Miller & Rollnick, 1991), led Miller to the development of an implicit model that guided his clinical behaviour. It was not until Miller was invited to lecture on behavioural interventions for alcohol problems and was forced to verbalise the underpinnings of these guidelines that the original clinical description of MI was formed (Miller, 1983, as cited in Miller & Rose, 2009).

What is Motivational Interviewing?

One of the beliefs of MI is that motivation to change lies within the individual and needs to be evoked in clients rather than imposed upon them. To adopt this approach, one must apply certain practitioner styles that guide the interviewing process (Miller & Rollnick, 2002). Specifically, the relationship between the practitioner and client must be collaborative rather than the practitioner taking on the role of an expert; the style of the practitioner needs to be one who searches for intrinsic motivation within clients and evokes it rather than giving advice or insight; the responsibility for change must be left with clients so it can arise from within them through a corresponding increase of intrinsic motivation.

This fundamental approach to MI briefly outlines the spirit of the method which has been identified as extremely important for its effective implementation (Miller & Rollnick, 2002). Without adopting this spirit (which will be described further in subsequent sections), a practitioner could succumb to what Miller and Rollnick (2002) call the “righting reflex”, or the desire to fix one’s clients. The authors specify that those working in helping professions may be particularly likely to demonstrate this righting reflex. Most individuals requiring a change in their lives exhibit “ambivalence” which is the normal human experience of feeling two ways about something. In the case of behaviour change, it is the feeling of wanting to, but also not wanting to, change. If someone who was ambivalent were to seek help from someone with a righting reflex, it is likely that the ambivalent individual would be offered advice, education, or persuasion as a resolution to his or her ambivalence (Miller & Rollnick, 2002). Due to his or her ambivalence, the most likely reaction of this individual would be to point out the flaws of the offered solution or argue against it. Miller and Rollnick emphasize the importance of inhibiting this righting reflex; the idea being that we as humans can talk ourselves into and out of things. Therefore through developing an argument against a solution imposed by a practitioner, an ambivalent individual can actually become less likely to change.

The model of MI combines these elements of practitioner style that constitute the spirit along with technique to focus on responding differently to types of client speech in the exploration and resolution of ambivalence (Miller, 1996; Miller & Rose, 2009). Responding to client’s speech requires attention to the natural language of change or “change talk”. Verbalizations of motivations for change, although potentially evoked by the practitioner, must come from the clients themselves. The practitioner then works to strengthen these verbalizations within his or her clients.

Counter change arguments or “sustain-talk”—when clients verbalise reasons not to or express unwillingness to change—can also occur. Miller and Rose (2009) specify that these kinds of verbalizations should be met with empathetic responses due to the knowledge that acting confrontationally would be counter-productive and could evoke defence of the current behaviour.

Both types of verbalisations (which will be described further in subsequent sections) represent different sides of ambivalence, described earlier. Although developing ambivalence—going from not wanting to change to wanting *and* not wanting to change—is part of the change process itself, it is also the most common place to get stuck along the change journey.

Motivational Interviewing Efficacy

Much of the initial work in MI was centred on substance abuse and other addictive behaviours, primarily because the development of the method began with the treatment of these individuals (Miller & Rollnick, 2002). However, MI has been widely extended to various behaviours such as increasing exercise, improving diet, reducing risky behaviour and improving engagement in further treatments. Several meta-analyses have been conducted to evaluate the efficacy of MI across this wide spectrum of behaviours; a recent meta-analysis covered 25 years of MI research and included 119 studies (Lundahl et al., 2010). The studies included in this meta-analysis covered a wide range of behaviours from substance abuse and gambling to eating problems and parenting practices.

It is evident from the meta-analysis’ small, yet significant effect size ($g = 0.22$) that MI is effective in producing significant change amongst a range of behaviours. However the results varied based on comparison group; MI was more effective when compared to a weak comparison group such as a waitlist control ($g = 0.28$) than when compared to a strong

comparison group such as a treatment with a specific programme ($g = 0.09$). Despite this variation, Lundahl et al. (2010) stated that, across the studies, 75% of MI participants gained at least some improvement, with 25% of these gaining at a moderate to strong level.

The use of MI has not been limited to adult populations. Research with youth has been relatively extensive, although the primary focus here remains on substance abuse behaviour. A recent literature review covering 39 of the most up-to-date studies of MI with adolescents in the substance abuse context saw mixed results (Barnett, Sussman, Smith, Rohrbach & Spruijt-Metz, 2012). However, the majority of studies (67%) reported statistically significant improvements in substance use outcomes including that of alcohol, tobacco, marijuana, and combinations of substances.

Additionally, MI youth research has extended to other clinical areas including the justice system. Stein, Monti et al. (2006) and Stein, Colby et al. (2006) investigated various outcomes of MI with a sample of substance abusing youth offenders. After an MI or Relaxation Therapy (RT) intervention, youth were mandated to a substance abuse treatment and their engagement was rated by the Treatment Participation Questionnaire (TPQ, Stein, Monti et al., 2006). Compared to RT, MI significantly reduced negative engagement such as joking around or making fun of others in treatment. Although not to a significant degree, MI also increased positive engagement such as appreciating the treatment staff and having aspirations about changing substance use or criminal behaviour. The MI group also had a significant reduction in driving under the influence of alcohol compared to control (Stein, Colby et al., 2006). Further research including the same youth sample saw those who received MI engaging in significantly lower rates of alcohol and marijuana use three months after treatment completion (Stein et al., 2011).

Overall, results are positive for MI across multiple behaviour domains; it is evident that MI is an effective intervention across various behaviours and populations. It was the extent of the research and new insights that prompted further development of MI and Miller and Rollnick's most recent revision of their book (Miller & Rollnick, 2012).

Further Development of Motivational Interviewing

This current MI book brought with it revised descriptions of its spirit and structure (Miller & Rollnick, 2012). Three levels of definition were also offered in replacement of prior approximations used to describe the method. The first focuses on the purpose of MI, a layperson's definition: "Motivational Interviewing is a collaborative conversation style for strengthening a person's own motivation and commitment to change." (pp. 12)

The spirit of Motivational Interviewing. The most recent description of the MI spirit contains four key elements: partnership, acceptance, compassion, and evocation. These interrelated components are not a replacement of the prior spirit, but rather a revision and extension. The central fundamentals remain consistent: emphasis is on the importance of a collaborative therapeutic relationship, motivations for change being evoked from the clients rather than imposed, and change being ultimately up to the clients and this responsibility left with them.

Partnership is an important factor in MI; it implies that the method is not done "to" or "on" someone, but "for" and "with" them. Rather than MI being something done by and expert to a passive recipient, there is a collaborative relationship between two experts; clients are recognised as experts on themselves. Through partnership rather than authority, this heightened respect can lead to a positive and supportive therapeutic relationship that is conducive to change. The main purpose of the practitioner in assisting with change is to

understand the lives of his or her clients through their eyes, rather than imposing a personal vision.

The attitude of client acceptance is related to the spirit of partnership in MI because it involves this profound respect. In accepting clients, it is not meant that their actions or behaviours are approved of. Acceptance involves: prizing the clients' absolute worth or having an unconditional positive regard for them (Rogers, 1980, as cited in Miller & Rollnick, 2012); demonstrating accurate empathy, an interest in understanding the internal perspective of clients without judgement; honouring and respecting their right and capacity of self-direction, their autonomy (Deci & Ryan, 1985; Markland, Ryan, Tobin & Rollnick, 2005, as cited in Miller & Rollnick, 2012); exhibiting affirmation or seeking and acknowledging the efforts and strengths of his or her clients, instead of searching for what is wrong and telling them how to fix it.

Compassion is the most recently added element of MI spirit. Here the focus is on giving priority to the needs of clients; compassion is a deliberate commitment to pursue the best interests of clients and their welfare. This element was added to the spirit of MI after the realization that all other components can be practised in pursuit of one's own interest. Therefore, by adding compassion it is ensured that the practitioner's intentions are directed at his or her clients.

Some models of therapy view clients as missing or lacking in something that needs to be installed or fixed. Unfortunately this approach does not usually work well with ambivalence. The final component of MI spirit is the strengths-focused premise of evocation: the belief that people have most of what they need already within them, it just needs to be called out. Evocation is related to the ambivalence concept mentioned above; ambivalent people want to change, but also do not want to at the same time. Arguments for change are

already there and the practitioner needs to evoke and strengthen these already present, positive motivations.

MI spirit is the context for the second definition from Miller and Rollnick (2012), that of the practitioner: “Motivational Interviewing is a person-centred counselling style for addressing the common problem of ambivalence about change” (pp. 21).

Additionally to the spirit there are four core communication skills central to the practice of MI. They create the acronym OARS and are derived from client centred counselling to help people explore ambivalence and identify change reasons (Miller & Rollnick, 2012). They are as follows: open-ended questions, these types of questions require an elaborative answer and encourage client speech and the facilitation of a trusting and accepting atmosphere; affirmation, statements of understanding and appreciation can be effective at building rapport and reinforcing exploring; reflective listening, a key skill in MI which makes the process of listening more conscious by decoding and verbalizing the client's thoughts and views back to them, this technique also assures the practitioner has an understanding of what was meant; and summarizing, a way the practitioner can show they have been listening by reflecting and pulling together what clients have said over a certain period of time or about a certain topic. Summarizing is also useful for reinforcing what has already been said and as preparation for further elaboration. An additional core skill that is not part of OARS is informing and advising. Giving advice or informing clients of something could influence their ideas about change in a certain direction, therefore a practitioner must only inform or advise his or her clients if they give permission or have asked.

The structure of Motivational Interviewing. Until the most recent MI text, its method was described as having two phases: building motivation for change and strengthening commitment to change (Miller & Rollnick, 2002). Recently it was

acknowledged that this original model seemed incomplete. A main issue was that in practice, the model failed to reflect a decision making process, one that is often seen as more circular than linear. The authors decided to reject these phases and instead describe MI as having four overlapping processes: engaging, focusing, evoking, and planning. These processes emerge in a particular order, however one does not end as the other begins; they essentially build upon one another. These four processes are described further by Miller and Rollnick (2012).

Engaging. Engaging clients in a collaborative working relationship is the first process of MI. Therapeutic engagement is a prerequisite for everything that happens in therapy, if there is no engagement, there is no progress. Positive engagement results in a higher likelihood of a client returning, without which therapy is ended. Another positive is that effective engagement builds a therapeutic alliance and the desired positive atmosphere that drives MI. This alliance was described by Miller and Rollnick (2012) to contain three aspects: a mutually respectful and trusting relationship, an agreement of the goals of treatment and the collaboration on the tasks required to reach the treatment goals.

During the engagement process, a major proportion of the practitioner responses should consist of reflective listening. Asking open-ended questions and reflecting during the process clarifies an understanding of a client and conveys this understanding to him or her. Here reflection demonstrates curiosity and interest in a client, promoting the collaborative therapeutic relationship.

Disengagement is undesirable and there are several traps that can lead therapy in that direction such as asking too many questions at a time or focusing the therapy too early. Asking a long series of questions sets the client's perception of the practitioner being in control, rather than the formulation of a partnership. Focusing the therapy before establishing the desired relationship could result in an inaccurate focus, perhaps on something the client is

not concerned about which could potentially lead to frustration. Other traps that could lead to disengagement include labelling, blaming, and engagement in small talk.

Effective engagement requires the practitioner to attend to issues that affect how a client may feel about the treatment. The practitioner needs to understand why the client has come to therapy, what he or she wants from the experience and how important his or her goals are. The atmosphere needs to be inviting so the client feels welcomed and respected. A sense of what to expect from therapy is important to make the client feel more knowledgeable and confident of what will occur. In explaining the therapeutic process, offering hope along with an honest positive picture of the efficacy of MI will help the client believe that he or she will achieve what is needed.

Focusing. The goal of focusing in MI is to answer the question “Why has the client come to therapy?” This is where the conversation about change is developed and maintained. The goals and direction of MI become clear at this stage as the practitioner and client determine what the client’s goals are and which should be targeted. A particular focus is formed through a conversation that has change at heart and can be from three possible sources: the client, sometimes a client will come to therapy with presenting concerns or problems; the setting, a particular agency may be funded to assist specific issues; the practitioner, it is common for clients to seek help with a goal in mind and the practitioner, through the consultation, identifies another required change.

It must be noted the importance of not hurrying the focusing phase. Rushing can lead to a premature focus that could possibly be inaccurate and undermine progress, even elongating the focusing process. The practitioner must show tolerance to uncertainty, an important component of skilful focusing in MI, and allow the process of focusing to take as much time as it needs.

Evoking. Once the practitioner and client have developed a working relationship and established a focus, the therapy turns distinctly MI with the addition of the third process of evoking and strengthening motivation to change. This process of evoking, previously known as eliciting change talk, has always been considered the heart of MI, and for that reason it is the basis for the final and most technical definition that explains how MI works:

“Motivational Interviewing is a collaborative, goal oriented style of communication with particular attention to the language of change. It is designed to strengthen personal motivation for and commitment to a specific goal by eliciting and exploring the person’s own reasons for change with an atmosphere of acceptance and compassion.” (pp. 29)

Several skills unique to MI are required throughout this process such as recognizing change talk and knowing how to evoke, strengthen and respond to it. MI practitioners also need to be efficient at recognizing sustain talk and responding to it appropriately.

Change talk. Change talk is described as “Any self-expressed language that is an argument for change” (Miller & Rollnick, 2012, pp. 159). Within this definition it became clear that several different speech acts were present and differentiation was required. There are two subgroups of change talk: preparatory change talk and mobilized change talk. Preparatory change talk is when clients verbalize the pro-change component of ambivalence and is made up of four types of speech: desire, a wanting to change; ability, acknowledging that change is possible; reasons, potential positive outcomes of change; and need, expressing the importance of change. Mobilized change talk comprises of language signalling a move towards change such as commitment, verbalizations like “I will” or “I promise”; activation, words indicating movement towards action, however without commitment; and taking steps, where the client indicates they have done something in the direction of change.

Evoking change talk. There is no one way of increasing client change talk, it is about listening and responding to each individual. If there is change talk, the same strategies that evoked it should continue. If sustain talk increases, another approach is required, Asking evocative questions is the simplest way to evoke change talk, essentially asking open-ended questions where change talk will be the answer. These questions should initially be related to preparatory change talk and when appropriate, questions related to mobilizing change talk can be included. It is common for mobilizing change talk to eventually occur after a period of evoking preparatory change talk.

Responding to change talk. There are particular ways to respond to change talk in MI. If a practitioner simply collects this talk, opportunities to consolidate change motivation will be missed. The four responses to change talk lie in the area of client centred counselling, such as the OARS techniques described earlier. Specifically, open-ended questions are used to ask more about the change talk, obtain examples and for elaboration. Affirmation is essentially praising the client for voicing change and can be as simple as a positive comment about a change statement. Reflective listening is a particularly effective response to change talk because when reflected, what the client says next is likely to be further change talk. Summaries are also effective as they pull together all the change talk that has been offered and deliver it back to the client, strategically directed towards change.

Although change talk highlights the importance of change, a client must have confidence in the change itself. This belief that change is possible is a very strong predictor of change and needs to be evoked in MI; commitment to change is unlikely without the confidence that change is possible. To strengthen this belief, confidence talk is the type of speech that needs to be evoked. The method of evoking this type of talk follows the familiar pattern of asking open-ended questions where answers are likely to be confidence talk and

then employing reflective listening. Responding to confidence talk is consistent with how the practitioner responds to change talk, employing the OARS techniques with reflective listening being the central skill.

Developing a discrepancy is also a motivational factor relevant to the evoking process of MI and can be seen as the distance between personal goals and the current state. Some clients come to therapy already conscious of a discrepancy within their lives; some even feel uneasy about it. There are three possible reasons why the knowledge of a discrepancy has not brought about natural change: the discrepancy may seem too large and daunting or too small and not sufficient to be important, there may be a lack of self-efficacy, or the discrepancy may evoke unpleasant feelings and the client has simply been avoiding it.

In MI, a discrepancy needs to be acknowledged in a way that does not cause self-defence within the client. The practitioner must create an atmosphere that is safe so a client can look at the uncomfortable truth and change from it. There needs to be acceptance rather than further discomfort; when clients accept themselves, change is possible. A way to reduce the negative feelings associated with a discrepancy is to reframe it from shame to opportunity.

Sustain talk. Verbalizations that are opposite to change talk are termed sustain talk, they argue against change. This language takes the same forms as change talk described earlier, just in a different context. Being the opposite to change talk, sustain talk predicts the client will not make a certain change.

Responding to sustain talk. Many practitioners believe that once change talk has been successfully evoked, the road is clear for change. However, sustain talk may arise throughout MI and it is important for this kind of speech to be met in an MI-consistent way, by following

the practices and spirit of the method; it needs to be understood that sustain talk is a naturally occurring part of ambivalence, not something pathological and oppositional.

There are several ways of responding to sustain talk; one is through the familiar use of reflective listening, although there are also strategic responses that can be effective with this kind of talk. These responses have the same intent as reflective ones, they are purely additional options: emphasizing autonomy, telling clients they have a choice and it is their choice heightens the possibility of change; reframing, suggest a different perspective or meaning for what a client said, thus providing another perception of the situation; agreeing with a twist, a reflection that sides with the client but with a slight reframe; running head start, getting a head start on the pros to change from listening to the cons, this method is trying to access change talk when none is present; coming alongside, used when yielding change talk has failed, this is agreeing with the client with no twist; sometimes joining the client's sustain talk triggers change talk.

Planning. Clients are more likely to follow through with change when they have a plan of action. This process of planning begins when a client reaches a certain stage of readiness and shows various signs such as increased change talk, less discussion of the problem, and envisioning the change. When the practitioner senses readiness for a change plan, the need is to determine whether the client is in fact ready. Asking directly is one option, another is recapitulation: a summary collecting up the change talk offered by the client.

This process is not a different form of MI, there is still change talk and the direction remains towards change, the same collaborative spirit is also employed; planning is an addition to the earlier processes. Change talk tends to be slightly different during planning; there is more mobilising change talk, verbalizations that indicate commitment, activation and

taking steps. It is however, not unusual to hear a mixture of mobilising and preparatory change talk; such a mix may be an indication of a low confidence level. Therefore language about confidence remains imperative at this stage, it is important the client believes in the plan. Part of this process is the development of a change plan addressing how change will fit with the client's life and how they will proceed with it.

There are three possible planning scenarios based on the client's goal. The simplest is when there is a clear plan, when there is only one way to achieve the goal and that plan is chosen and pursued. Sometimes there are various clear options, they need to be discussed and one chosen. There are also instances where a goal is clear, but how to achieve it is unclear, here a plan needs to be developed from scratch.

These three scenarios can send therapy on slightly different roads to the formulation of a change plan. However, once one has been selected, a practitioner needs to summarize it to make sure both his or her clients and he or she understand and agree on it. After the summary, a practitioner would then use open-ended questions that elicit mobilizing change talk and commitment. A useful strategy when formulating a plan is to troubleshoot it. Here the practitioner asks the client about what could go wrong with the plan, client answers will most likely be mobilizing change talk; this strategy also helps in formulating solutions to potential problems with a plan.

When a change plan has been developed, there is a further step from planning to action. It is important to determine whether the client is happy with the plan and is intending to carry it out. By listening for mobilizing language, a practitioner can identify what clients are willing and intending to do. A way of strengthening commitment could be to arrange for a client to agree to particular steps within the change plan rather than committing to the overall goal straight away. By asking clients about specific steps they are intending to take during the

next week as opposed to whether they are intending on committing to the whole plan puts less pressure on clients and therapeutic relationships. This approach zooms in on a specific component of the plan, viewing a step that is doable rather than a long-term, potentially intimidating process.

Committing to change effectively ends the MI process as a whole and some clients need little or no assistance from this point. It must be noted that committing to change could occur prior to the planning stage; some clients feel comfortable enough to pursue with change without making a plan or continuing with therapy. However, there are other clients that prefer the continued support throughout the change process. Continuing the spirit and style of MI alongside other tools and skills can help to facilitate the change process.

Miller and Rollnick (2012) express that MI does not provide a comprehensive change model. MI is more of a tool to move past ambivalence and get people on the road to change. Early findings showed that once clients had successfully progressed through the evoking and or planning stages, with revisits if necessary, they proceed with change on their own; MI gets them past the hardest task of deciding to make the change.

The revision of the spirit and structure of MI has created a more complete model with which one can follow when learning the method. The current MI spirit with its four key elements has extended upon earlier descriptions and particularly emphasizes further the importance of client's welfare and best interests through the addition of the element of compassion. The four processes rather than two phases help to guide the structure of therapy without being over instructive; by approximating what one should encounter when practicing MI as opposed to strict sequential guidelines, therapy can take shape depending on the client or the change at hand.

It is clear that MI is a highly developed method with a significant amount of outcome research to support its efficacy across many contexts. However, like other evidence-based interventions, efficacy relies on the training of the practitioners implementing the method. Without sufficient training, effective implementation can be compromised, adversely affecting outcomes. Research on implementation training has made specifications to what defines a well-operationalized programme. These specifications are highly important to effective training programmes, thus relevant to the current research.

Implementation Training

It must be acknowledged that training does not guarantee that implementation of an intervention will be at a satisfactory level. Adequate programmes can, however, increase the quality of which an intervention is delivered through integrity improvements and increased consistency across practitioner behaviour (Aarons et al., 2011; Feldstein et al., 2008; Fixsen et al., 2005; Stirman et al., 2004, as cited in Becker & Stirman, 2011).

Members of the National Implementation Research Network, Blase and Fixsen (2013) indicate that, for an implementation programme to be well-operationalized, core components (elements key to the success of the intervention) need to be identified. Also important is the knowledge of what elements can be modified without adversely affecting outcomes. Because evidence-based implementation programmes such as that of MI are constantly being replicated, there should be emphasis on what is required for consistent positive outcomes.

Specifically, core components are elements of an implementation programme necessary for the production of desired outcomes (Blase & Fixsen, 2013). They intend to impact the proximal training outcomes in a positive manner and therefore increase the likelihood of positive outcomes long-term. Core components can take several forms within the intervention context such as being theory-based, derived from empirical research;

contextual factors, relating to the intervention from the environment; or structural elements, for example the required number of sessions for the intervention and specific intervention practices that are essential for successful implementation.

Identifying core components can be beneficial when evaluating a given programme, determining whether it is effective or needs improvement. If a programme is not proving effective, by knowing these essential elements, less time would be wasted trying to specify what is needed to improve outcomes; the focus would be on whether the core components have been implemented efficiently or if there are other factors inhibiting programme success. Several additional benefits are also offered by Blase and Fixsen (2013) and are as follows:

- Core components would provide the knowledge of what variables resources should be directed at for the production of the most desirable outcome.
- The likelihood of interpreting outcomes accurately and understanding what needs to be addressed for improvement would be increased. Without the understanding of core components, it is difficult to distinguish between an implementation and an effectiveness problem.
- Replications would be more feasible with less risk of programme ‘drift’, where unintentional changes result from replication, poorly affecting results.
- The ability to build theory that produces positive outcomes consistently across diverse settings or participants.

These authors also state that if an implementation programme is to be useful in everyday settings it must meet certain criterion that includes core components. This criterion is as follows:

- The context of the programme needs to be clearly described. This includes both a definition of the intended population, but also values and principles that underlie the programme and help to guide implementation when a scenario is complex.
- A description of the core components: what constitutes these components along with a range of how identification of these essential elements can benefit an implementation programme (described earlier).
- Practice profiles need to be established to specify aspects of the core components to a point that they are learnable and teachable.
- The practitioners who run the programme must undergo practical assessments to ensure they are engaging in behaviours that reflect the core components. From these assessments, measures of the extent to which core components are being adhered to, can be obtained.

It is expected that by following the above criterion, a given programme would be well-operationalized (Blase & Fixsen, 2013). Focusing research on core components is likely to improve the understanding of the necessities of evidence-based programmes, essentially being able to identify ‘what works’.

The Training of Motivational Interviewing

Although the various guidelines of MI training were published prior to that of Blase and Fixsen’s (2013) criteria for a well-operationalized programme, certain aspects of the most recent guidelines (Miller & Rollnick, 2012) can be viewed as consistent with these strategies. Miller and Rollnick (2012) suggest the content of any MI training programme to consist of four broad elements that stay close to what the research has been identified to be important: MI knowledge and spirit, introducing the spirit and processes of MI through a

workshop that increases knowledge and includes demonstrations can provide a foundation to be built upon with further training; the development of proficiency in the OARS client centred counselling skills is also necessary as they are fundamental to the practice of MI; learners then need to develop the skills of focusing along with recognising, evoking, responding to and strengthening change talk. The final component concerns learning the process of planning; here several skills such as timing, developing a change plan and evoking change commitment need to be mastered. It is recommended that, as a minimum MI training should consist of these four elements.

Turning back to the research of Blase and Fixsen (2013), these guidelines proposed by Miller and Rollnick (2012) can be viewed to reflect theory-based core components. Being derived from research, the four broad components identify the necessary MI content; the minimum requirement for an effective training programme, regardless of context. Although much more can be added to a given training programme, these components are central to MI and facilitators will know to dedicate sufficient time to ensuring their learners understand these areas.

There is however, no standardised training format for MI. Miller and Rollnick (2002) believe that having specific programmes for teaching the method, where an expert provides information and answers to passive participants, was too similar to the expert model of therapy that MI sought to avoid. There are however, guidelines described by Miller and Rollnick that are a result of outcome research in MI training. Whilst these authors acknowledge there are many different styles of which the learning of MI can be facilitated, the guidelines have been viewed as effective and could perhaps develop into further core components of MI training.

The first guideline is in regards to training in the form of a workshop or self-study. Although effective at increasing knowledge in MI, Miller and Rollnick (2012) state that this type of training alone is unlikely to improve skill and competence. Evidence supporting this claim comes from Miller, Yahne, Moyers, Martinez, and Pirritano (2004) who conducted a randomised controlled trial of different MI training methods. All groups apart from a wait-list control received a two-day workshop and submitted in-session audios. Their progress, reflected in the Motivational Interviewing Skills Code (MISC), was assessed at various follow-up times. The condition that received solely the two-day workshop, although initially demonstrating a significant increase in MI proficiency immediately following the training, failed to achieve clinical trial proficiency throughout the four and eight month follow-ups. This group fell to a level that was comparable to the wait-list control condition who had received only the practitioner manual and training videos. Without further training, this group was unable to maintain proficiency in MI. Miller and Rollnick emphasise that it is unlikely practical skill will result from workshops; however they can be used as initial training introducing the method.

Along with initial workshops, follow-up practical training is required for improving MI proficiency. Feedback is important here, Miller and Rollnick (2012) recommend that feedback and coaching based on observed practice is essential which is best done through in-session audio recordings. Measures such as the MISC and the Motivational Interviewing Treatment Integrity Code (MITI) can be used to assess MI behaviour from session recordings. Miller et al. (2004) included three 'enhanced training' conditions in their research. All three conditions received the workshop along with an additional component: feedback on direct observation of in-session audios, via email or a standard reporting form, six individual telephone coaching sessions that were half an hour long, or both of the previously described components. On average, these conditions met the standard for clinical

trial proficiency in MI, along with the proficiency standard for MI-consistent responses to clients (95%) at the four and eight month follow-ups, measured by the MISC. Miller and Rollnick acknowledge that individual practitioners may require differing amounts and levels of coaching depending on how readily they become proficient in MI. The authors recommend on-going direct observation of practice to maintain the skills learnt in MI. Skills can drift overtime and the use of coding measures can determine whether practitioners are constantly meeting a level of proficiency.

Evaluating Motivational Interviewing Training

To understand whether a given training programme is effective, the skill attainment of the participants must be evaluated. In assessing this, validated reliable measures must be used. The goal of the present study is to investigate whether two under researched MI measures are reliable and valid and could potentially make evaluating MI training less labour-intensive. Therefore a review of all the relevant measures is warranted. Each measure is thoroughly described along with accounts of previous psychometric evaluations. Additionally, summaries of training outcome research are included and give insight into the extent to which each measure has been utilised across the MI research, particularly how little the written-based measures appear in this literature.

Measures evaluating in-session practice. The most frequently implemented and valid measures of MI skill attainment are the MISC and the MITI which, as mentioned earlier, evaluate MI practice and adherence from in-session recordings. Across the MI literature, both the MISC and the MITI have been utilised across the common methods of MI training: workshop and workshop along with feedback and coaching. However, these measures have appeared more commonly across evaluations of trainings that included the feedback and caching. The MISC and the MITI are described in detail below along with

psychometric information and summaries of training outcome research. The summaries are arranged in a way that brief overviews of the studies precede the outcomes.

The MISC. Through in-session recordings the MISC assesses MI practice by rating and quantifying both practitioner and client responses. The most recent version (MISC 2.1) requires a coder to listen to each recording three times, rating different aspects of practice each time (Miller, Moyers, Ernst & Amrhein, 2008). When listening the first time, the session is played uninterrupted and the coder completes a set of global scales. The global scores are designed to reflect the coder's overall impression of the practitioner performance and client responses. The coder must assign a number from a seven-point Likert scale for each of the ratings to characterise the interaction. For the practitioner, the three dimensions of acceptance, empathy, and spirit are rated; brief descriptions are as follows (Miller et al., 2008):

- **Acceptance:** This rating intends to measure the extent to which the practitioner is respectful and holds a positive regard for the client. A practitioner who demonstrates high acceptance can be perceived as warm and supportive, respecting client's decisions without agreeing or approving of them. A practitioner low in acceptance may be perceived as disrespectful or judgemental.
- **Empathy:** The intent of this rating is to measure the extent to which the practitioner makes an effort to understand the client's views, typically through reflective listening. A practitioner with high empathy makes an effort to ensure his or her clients are understood. A practitioner with low empathy shows little interest in understanding his or her clients and generally keeps the conversation about factual information.
- **Spirit:** This rating intends to measure practitioner competence in using MI. As the spirit is central to the practice of the method, the core characteristics of

collaboration, evocation and autonomy (described earlier) are focussed on in this rating. The difference between practitioners with high or low MI spirit lies in their demonstration of these characteristics.

For clients, a single rating reflecting the period of most self-exploration is completed. Again made on a seven-point Likert scale, a score is given on the basis of the amount of personally relevant material shared in the session, along with the extent to which feelings, values, perspectives and perceptions were shared.

When listening to the recording a second time, behaviour classifications are completed for all practitioner utterances. An utterance is defined as a complete thought and is terminated by a new idea or a client response. For the practitioner, each utterance is assigned to one of the following 15 behaviour categories: advise, affirm, confront, direct, emphasize, control, facilitate, giving information, question, raise concern, reflect, reframe, support, structure, and warn. Further descriptions of the behaviour categories are detailed in Miller et al. (2008).

Four of these categories divide further into subcategories and are as follows: advise and raise concern are divided into whether or not advice or concern was given with permission, questions are divided into open-ended or closed questions (will be described further in subsequent sections), and reflect is divided into simple or complex reflections (also will be described further in subsequent sections).

Finally, when listening to the recording the third time, the type, intensity and frequency of client language are recorded. The first task is to clearly define the target behaviour change (TBC); understanding the TBC helps define the client utterances as being towards or away from behaviour change. There are five content codes that can be assigned to client utterances: commitment, reason, taking steps, follow/neutral, and other. The reason

code has three sub codes: desire, ability, and need, therefore if an utterance was coded as reason the coder must decide a sub code to be assigned also. With the exception of follow/neutral, each coded utterance must be assigned either a positive (+) or negative (-) sign to indicate whether the client is expressing change talk or sustain talk. These speech forms are described further in Miller et al. (2008). When an utterance is not able to be coded by the above speech forms as it does not reflect movement towards or away from change, it is disregarded. Only talk regarding the TBC can be coded.

The strength of client utterances may also be assigned, however this component is optional. A given utterance can be rated as high, medium or low and as there are no natural categories of language intensity, it was acknowledged these ratings may be less precise than other parts of the MISC. Examples of the utterance strengths are described in Miller et al. (2008).

Summary scores can be produced using the behaviour counts from the MISC (Miller et al., 2008). Of the questions from a given session, the percentage of open-ended questions (%OQ) can be calculated. The same can be done with reflections to find the percentage of complex reflections (%CR). A reflection to question ratio (R: Q) can also be generated. By summing the amount of responses coded as advising with permission, affirm, emphasize control, open-ended question, reflect, reframe, and support the amount of MI consistent responses (MI-Con) can be computed. The same can be done for the MI inconsistent responses (MI-In) by summing the utterances coded as advising without permission, confront, direct, raise concern without permission, and warn. The percentages of these responses can also be computed (%MI-Con and %MI-In). Finally, the results of the MISC can also be used to compute the percentage of client change talk (%CCT).

Proficiency standards using the summary scores have remained consistent across versions of the MISC and were reported by Baer et al. (2004). Firstly, the standard for global scores is above 5 on the Likert scale. To achieve beginning proficiency, 95% of clinician behaviour must be MI consistent, 50% of questions must be open-ended, 40% of reflections must be complex, and the R: Q must exceed 1:1.

Psychometrics. It must be noted that psychometric evaluation was only found for the older version of the MISC (1.0); it appears no such investigations as of yet have been made into the MISC 2.1. The MISC 1.0 was less refined and included more global rating scales.

Reliability of the MISC has been evaluated in several studies (Baer et al., 2004; de Jonge, Schippers & Schaap, 2005; Moyers, Martin, Catley, Harris & Ahluwalia, 2003; Tappin et al., 2000). Tappin et al. assessed inter-rater reliability of the MISC by categorizing the global items to form global practitioner ratings, global client ratings and interaction ratings. The intra-class correlations (ICCs) were small to moderate at (.39, .53, and .51 respectively). The behaviour counts were also categorized to form practitioner behaviour consistent with MI, practitioner behaviour inconsistent with MI, client behaviour consistent with MI and client behaviour inconsistent with MI. ICCs were higher for these categories (.45, .67, .77, and .76 respectively).

ICCs for the five summary scores of the MISC were computed by Baer et al. (2004). All ICCs were in the good to excellent range (.66–.95) apart from %MI-Con and %MI-In (both -.17). The authors omitted the %MI-In summary score and modified the %MI-Con score to reflect the percentage when including both neutral and MI inconsistent utterances; the corresponding inter-rater reliability was strong (.84).

Moyers et al. (2003) and de Jonge et al. (2005) investigated each item of the MISC. Global ratings differed across the studies with Moyers et al. reporting a range of poor to

excellent ICCs (.252–.791) whereas de Jonge et al. obtained only poor ICCs (.06–.44). The latter authors also used the Gower coefficient of agreement (which will be described further in subsequent sections) between pairs of coders to obtain the absolute agreement for each global rating item. These coefficients ranged between .77 and .94 indicating there was 77–94% similarity in rating across pairs.

The same range of ICCs was reported by Moyers et al. (2003) for the behaviour counts (.000–1.000). De Jonge et al. (2005) used chi square analyses for these items where they identified several differences across raters on various items including affirm, confront, and support.

These authors also evaluated content validity of the MISC through coding of a transcript with known MI consistent and MI inconsistent utterances. Although all the MI principles were identified, they were unbalanced. De Jonge et al. (2005) noted that express empathy was covered by the majority of the global rating items whereas developing discrepancy, roll with resistance, increase self-efficacy, and avoid argumentation were found among the behaviour counts. Also the traps to avoid (outlined by Miller & Rollnick, 2002) were not clearly covered by the MISC. This was identified when “falling into the expert trap”—when a practitioner communicates that they are in charge and sets up the expectation that he or she will have the answers to the client’s problems— was present twice within a transcript, however not recorded by the measure.

Psychometric evaluations of the MISC indicated that the measure demonstrated low to good reliability. The variance in reliabilities paired with the unbalanced representation of the MI principles may have led to refinement and the development of the MISC 2.0 and 2.1, and eventually the MITI.

Literature Summary. The amount of research using the MISC is relatively modest compared to the MITI, however this unbalance is understandable given the MITI effectively replaced the MISC as a refined improvement; the most likely reason the MISC would be chosen over the MITI in more recent research would be if client responses were of interest along with that of the practitioner.

Five studies included the MISC in their training evaluations across different populations including substance abuse counsellors (Baer et al., 2004; Miller et al., 2004), diabetes nurse educators (Britt & Blampied, 2010), dieticians (Brug et al., 2007), and mental health therapists (Schoener, Madeja, Henderson, Ondersma & Janisse, 2006). The majority of studies evaluated their curriculums using a single group of participants (Baer et al., 2004; Britt & Blampied, 2010; Schoener et al., 2006), however Brug et al. (2007) and Miller et al. (2004) used randomised controlled trials (RCT). Brug et al. randomly divided dieticians to either receive or not receive the training, whereas Miller et al. divided substance abuse counsellors into five different training groups in order to explore the variability in outcome of different types of training, described earlier.

Due to the quantity of variables measured by the MISC, not all studies investigated every possible outcome. The following are grouped and reported according to each MISC scale, resulting in individual summaries for each outcome.

Of the studies that reported the MI spirit rating, Miller et al. (2004) and Schoener et al. (2006) observed significant increases from pre- to post-training. Miller et al. also reported differences among their training groups that begun at the four-month follow-up. Results indicated the three groups who received feedback or coaching were meeting the proficiency standard for this rating, whereas the workshop and self-study groups were not. Brug et al. (2007) did not observe a significant difference in MI spirit within a month after post-training

compared to control. However, at the five to six month follow-up, when the feedback sessions were completed, the dieticians who received the training performed significantly better on the rating.

The empathy rating was reported by Brug et al. (2007) and Schoener et al. (2006) with results consistent with those of the spirit rating. Schoener et al. observed a significant increase in empathy as an effect of training, and although not significant at the first post-training assessment, the training group in Brug et al. performed significantly better than the control at the second assessment.

Several studies included the behavioural counts regarding %OQ (Baer et al., 2004; Britt & Blampied, 2010; Brug et al., 2007; Schoener et al., 2006). Baer et al. (2004) and Schoener et al. (2006) reported significant increases in this rating from pre- to post-training with Schoener et al. also reporting a significant decrease in closed questions. The increase reported by Baer et al. however, did not hold across follow-up, returning to a level comparable with baseline. Britt and Blampied (2010), whose sample size was too small for significance testing, reported results with regards to the proficiency standards (Baer et al., 2004). At post-training, the diabetes nurse educators achieved a level of %OQ approximating beginning proficiency. The open along with the closed-question ratings were also reported by Brug et al. (2007), however were not significantly better than control at both assessments.

Results of the reflection counts also varied across different curriculums. The %CR was reported by Britt and Blampied (2010) where an increase that exceeded the level required for competency (50%) was observed at post-training. Schoener et al. (2006) also reported a positive result, a significant increase from pre- to post-training in the use of both simple and complex reflections. Simple and complex reflections along with the total use of reflections were reported by Brug et al. (2007) with mixed results. At the first assessment neither simple

nor complex reflections significantly differed between training and control group. At the second test the training group used significantly more simple reflections than control; however the difference in %CR remained not significant. When totalling the reflections, the training group used significantly more at both tests than at control. Finally Baer et al. (2004) reported %CR at pre- and post-training. Although the substance abuse counsellors used a higher %CR than simple reflections, the percentage did not significantly increase from pre- to post-training or across follow-up.

The R: Q exhibited more consistent results across the studies. A significant increase was observed by Baer et al. (2004) as a result of training which remained consistent and significantly higher than baseline at follow-up. Similar results were reported by Britt and Blampied (2010) where the increase in R: Q exceeded the level required for beginning proficiency (1:1) and in the case of one participant, reached the level of competence (2:1).

The %MI-Con was computed in three studies (Baer et al., 2004; Britt & Blampied, 2010; Miller et al., 2004), with consistent results. All studies observed an increase from pre- to post-training on this rating. Specifically Britt and Blampied (2010) reported an increase to levels comparable with beginning proficiency (95%), and although a significant increase was observed by Baer et al. (2004), %MI-Con reduced during follow-up to a level comparable with baseline. Again Miller et al. (2004) reported differences among the training groups. Results were similar to that of the MI spirit rating: only the three enhanced training groups met the proficiency standard at the four-month follow-up.

Finally, as the MISC also takes into account client variables, %CCT was also reported across the research. All studies reported an increase in %CCT statements with the exception of Baer et al. (2004) who reported the percentage of clinician talk time which did not significantly decrease across training. Specifically, Britt and Blampied (2010) and Schoener

et al. (2006) reported a larger %CCT as an effect of training. Again, with regards to Brug et al. (2007), percentages were comparable across the training and control group at the first assessment; however clients in the training group exhibited significantly more change talk at the second. Although there was an increase in %CCT from pre- to post-training, Miller et al. (2004) reported that at the four-month follow-up, only one of the enhanced training groups (workshop plus feedback and coaching) exhibited significantly better client responses compared to baseline.

Along with Britt and Blampied (2010), Baer et al. (2004) also compared their results to the proficiency standards, outlined above. Results indicated that beginning proficiency was met for the %CR (40%) across pre- and post-training along with follow-up. The R: Q increased from pre- to post-training to meet beginning proficiency, however decreased to below the threshold by follow-up.

Comparisons with proficiency standards were not reported in the research by Brug et al. (2007) and Miller et al. (2004). However, it was possible to compare the results from the studies to the levels of proficiency. Solely the global scores could be interpreted this way in the research by Brug et al. Results indicated that at neither time point did the trained participants reach proficiency in MI spirit or empathy. With regards to the research of Miller et al., at the four-month follow-up results varied across the summary scores. Only the groups who received enhanced MI training were proficient in MI spirit, the %MI-Con, and %OQ. All groups with the exception of the self-training control were proficient in the R: Q and all groups regardless of condition were proficient in the %CR.

Because the MISC has been demonstrated to have relatively good reliability, the variation in results across the research was thought to reflect the different training designs; when taking into account the differences among trainings, a familiar pattern emerged. All studies with the exception of Baer et al. (2004) included feedback and or coaching

components in their trainings, resulting in predominantly positive outcomes. Britt and Blampied (2010) and Schoener et al. (2006) observed exclusively positive results from their curriculums. Miller et al. (2004) reported stronger MISC results from the groups who received feedback and or coaching as part of their training and, although Brug et al. (2007) reported a large amount of results that were not significant at the first post-training assessment, at the second when the feedback component of training had been completed, several MISC ratings reached significance. The outcomes of Baer et al. appeared typical of a curriculum that consisted of a theoretical component such as a workshop with no additional feedback or coaching: initially positive gains were made immediately post-training, however the majority of results reversed across follow-up. These particular patterns of results strengthen the validity of the MISC by being consistent with the training recommendations of Miller and Rollnick (2012) who emphasize the importance of feedback and coaching to strengthen the MI skills learnt during training.

The MITI. The MITI was derived from the MISC; it was decided that a more reliable and economical assessment of practitioner adherence to MI was required (Moyers, Martin, Manuel, Hendrickson & Miller, 2005). The MISC is labour intensive with the requirement of listening to in-session recordings multiple times. The measure also examines client variables along with that of the practitioner. Therefore an exploratory factor analysis was performed condensing the MISC to reduce the complexity and number of ratings, thus creating the MITI.

Two components make up the MITI, global scores and behaviour counts; both assessed during a single review, which is usually a randomly selected 20-minute session segment (Moyers, Martin, Manuel, Miller & Ernst, 2010). Global scores capture the overall coder judgement of the session with regards to practitioner performance, covered by five

dimensions: evocation, collaboration, autonomy/support, direction, and empathy. A single number from a five-point scale is assigned to each dimension to characterise the whole interaction. There are summaries for each rating that specifies what constitutes a certain score, brief descriptions are as follows (Moyers et al., 2010):

- **Evocation** – Measures the practitioner's understanding of clients' motivation for and ability to change being within them, thus focussing on eliciting these motivations and abilities. A low score on this scale indicates opportunities to explore ambivalence are generally missed due to the practitioner's interest lacking in this important area. A high score reflects a practitioner who explores clients' personal change reasons and is curious about their ideas of change.
- **Collaboration** – Determines whether the practitioner sees the therapeutic relationship as collaborative: to what extent their behaviour demonstrates that the interview is between two equal partners. A low collaboration score reflects a practitioner using his or her authority to make progress and does not make an effort to develop a mutual understanding of the goals of treatment. A high score indicates the practitioner is working with the client co-operatively.
- **Autonomy/Support** – Measures the level of support and fostering of the client's perception of choice provided by the practitioner. A low score indicates the practitioner believes that clients cannot change their behaviour without input, therefore attempts to control the client's choices. A high score reflects a practitioner who believes that clients have the potential to move in the direction of change and helps them to acknowledge choices regarding the target behaviour.

- Direction – Does not always reflect the use of MI as the concern is with the practitioner maintaining an appropriate focus throughout a session. A low score indicates the session lacks appropriate aims and structure along with a lack of exploration of behaviour change. A high score indicates a good focus on the target behaviour and immense effort to maintain this appropriate focus. However, a domineering practitioner can score highly on direction, emphasizing that a high score on this scale does not indicate better use of MI.
- Empathy – Determines the extent to which a practitioner makes an effort to understand a client's feelings about or perceptions of change. A low score indicates indifference about client perceptions with little effort to understand these feelings; this kind of practitioner might act in a hostile manner with a client. A high score reflects a practitioner who is curious about his or her clients and explores their perceptions and feelings of change.

Behaviour counts record instances of practitioner utterances, reflecting particular behaviours, using a running tally throughout the reviewed session segment. Unlike the global scores, the quality of the behaviour is not judged but simply counted. Only one behaviour code can be assigned to each utterance. There are five primary behaviour codes (Moyers et al., 2010): giving information, questions, reflections, MI-adherent behaviour, and MI non-adherent behaviour.

An utterance is termed giving information when the practitioner appears to be educating their clients, providing feedback or disclosing personal information to them. This code is not chosen when an opinion is given without advising and should not be confused with the MI non-adherent behaviours of confronting, directing or giving advice.

With regards to questions, secondary codes fall into this category:

- Closed Questions: can be answered with a yes, no or unelaborated response.
- Open-ended Questions: can be answered with a wide range of possible responses and intend to seek information and encourage exploration.
- Questions trying to be reflections: a statement meeting the criteria for a reflection, however sounding like a question. These statements must be coded as either open-ended or closed questions.

The intention of the reflections category is to identify reflective listening statements made in response to a client. There are also secondary codes:

- Simple Reflections: where the aim is to convey an understanding or attempt to facilitate practitioner and client exchanges; little or no meaning is added to what the client has said.
- Complex Reflections: a deeper, more complex picture is conveyed by adding emphasis or meaning to what the client has said.

An utterance is termed MI-adherent when a practitioner's behaviour is consistent with an MI approach such as asking permission before giving advice or information, affirming, emphasizing client control and supporting the client with sympathy or compassion. If the utterance clearly falls into the MI Adherent category, it should be coded as such, however if in doubt, a different code such as a question or reflection should be given. Statements that open or close the session are formalities and are not coded unless specific MI Adherent behaviour is identified.

An utterance is termed MI non-adherent when a practitioner's behaviour is inconsistent with an MI approach such as advising without permission, confronting and directing. Analogous to the MI Adherent category, if an utterance clearly falls into the MI

non-adherent category, it should be coded as such; however a different code should be given if in doubt.

It must be noted that there are also utterances not coded by the MITI. These include greetings, structure statements, self-disclosure statements, previous session content, incomplete thoughts, and material that are off topic.

Summary scores for the MITI were developed and can be interpreted using practitioner proficiency and competency thresholds (Moyers et al., 2010). Global spirit ratings can be calculated by adding the scores from the dimensions of Evocation, Collaboration and Autonomy/Support and dividing the answer by three. Percentages of complex reflections (%CR), open-ended questions (%OQ), MI-adherent behaviour (%MI-Ad), along with a reflection-to-question ratio (R: Q) are also summary scores. A coding sheet is used to record summary scores along with the behaviour counts (see Appendix A for the MITI coding sheet).

According to the thresholds derived by Moyers et al. (2010), to meet MI beginning proficiency, the global spirit rating must exceed 3.5, the R: Q must exceed one, the %OQ must exceed 50%, %CR must exceed 40%, and %MI-Ad must exceed 90%. To meet the level of competency, the global spirit score must exceed four, R: Q must exceed two, %OQ must exceed 70%, %CR must exceed 50% and behaviour should be 100% MI-adherent. It must be noted here that these thresholds are based on expert opinion and lack supportive psychometric data (Moyers et al., 2010).

Psychometrics. The psychometrics for the MITI were obtained through a relatively large study of practitioners enrolled in a randomised controlled trial investigating various MI learning stages (Miller et al., in press, as cited in Moyers et al., 2005). Participants were asked to submit audios with a standardised patient portraying the role of a client before and

immediately after training and at four, eight, and 12 months post training. Randomly selected 20-minute session segment tapes were coded by two university undergraduate students and one graduate.

Inter-rater reliability was estimated from 50 randomly selected tapes that were then coded independently by all three coders. ICCs varied across global ratings and behaviour counts; however they were all greater than .50 with 70% in the excellent range, derived from the guidelines of Cicchetti (1994). Cronbach's alpha (α) coefficients were also calculated and ranged from .76 to .95 demonstrating good to excellent internal consistency. Convergent validity with the MISC was established through canonical correlations between the factor analysis scores (conducted for the MITI development) and MITI variables. Six of these correlations were significant which accounted for 59% of the variance in the MISC factor scores.

Further analysis of the psychometric properties of the MITI was carried out by Pierson et al. (2007). In this research, practitioner's audios or role-plays of both 10- and 20-minute session segments were coded to determine length comparability. The coders in this study had little or no experience with therapeutic interventions.

Even under these highly different conditions (compared to Moyers et al., 2005) ICCs ranged from fair to excellent ($>.40$), indicating the MITI can be efficiently coded by raters with little or no experience with MI. Pierson correlations were calculated to assess the validity of the MITI and determined that the measure was in accord with MI theory. In regards to the audio segment length, paired-samples t-tests were carried out to determine if the 10- and 20-minute session audios were comparable. The two lengths differed in the rates of some behaviour such as MI-Ad and %CR. However, Spearman correlations showed high

consistency for each item from the two differing segments; therefore 10-minute segments were acknowledged to be useful for reliably measuring MI adherence.

Literature summary. The MITI has been used extensively throughout the literature. Sixteen studies included the MITI for evaluation of training across several populations including nurses (Bohman, Forsberg, Ghaderi & Rasmussen, 2013; El-Mallakh, Chlebowy, Wall, Myers & Cloud, 2012; Robbins, Pfeiffer, Maier, LaDrig & Berg-Smith, 2011), medical students (Daeppen et al., 2012), AIDS counsellors (Evangeli et al., 2009), substance abuse clinicians (Mitcheson, Bhavsar & McCambridge., 2009; Smith et al., 2007; Smith et al., 2012), smoking counsellors (Forsberg, Forsberg, Lindqvist & Helgason, 2010), behavioural health providers (Moyers et al., 2008), health care professionals (van Eijk-Hustings, Daemen, Schaper & Vrijhoef, 2011; Wahab, Menon & Szalacha, 2008; Wolfe et al., 2013), trainee counsellors (Young & Hagedorn, 2011), prison staff, and probation officers (Forsberg, Ernst & Farbring, 2011; Hartzler, Baer, Dunn, Rosengren & Wells, 2007; Walters, Vader, Nguyen, Harris & Eells, 2010).

As with the MISC, not all studies reported every possible outcome on the MITI. Due to the magnitude of research utilising this measure, MI skill was evaluated across several outcomes: improvements from pre- to post-training, improvements across coaching and feedback components of training, and differences between trained and untrained (control) participants. The following report organised the research according to these three outcomes.

Three studies evaluated their training outcomes by taking MITI measurements at pre- and post-training (Evangeli et al., 2009; Hartzler et al., 2007; Wolfe et al., 2013). Hartzler et al. (2007) and Wolfe et al. (2013) included feedback and coaching components in their research whereas Evangeli et al. (2009) solely included workshops. All studies reported the global scores of MI spirit and empathy. Evangeli et al. and Hartzler et al. reported the %OQ,

%CR, and %MI-Ad. Solely Evangelini et al. reported the R: Q, whereas Wolfe et al. did not report summary scores,

Hartzler et al. (2007) reported a significant increase in MI spirit from pre- to post-training; however the increase in empathy was not significant. Evangelini et al. (2009) reported MI spirit and empathy as one outcome (named global therapist ratings) which did not yield a significant increase. Due to their number of participants, Wolfe et al. (2013) did not carry out significance testing. Increases were reported on both global MITI outcomes.

With regards to the behaviour counts, both Evangelini et al. (2009) and Hartzler et al. (2007) did not observe significant increases in the %OQ and %CR; Evangelini et al. however reported a significant increase in the R: Q. There was inconsistency across the studies for %MI-Ad; Evangelini et al. observed a significant increase whereas Hartzler et al. did not.

Evangelini et al. (2009) and Wolfe et al. (2013) compared the results of their participants to the proficiency thresholds derived by Moyers et al. (2010). Both studies reported the number of participants who achieved beginning proficiency on each of the MITI outcomes. At post-training, all four participants from Wolfe et al. achieved proficiency on MI spirit; three achieved this level on empathy, and three on direction. Of the 17 participants evaluated by Evangelini et al., four achieved beginning proficiency on the global therapist ratings, nine achieved this standard for %OQ, one for %CR, zero for the R: Q, and three on MI-Ad.

Additionally Evangelini et al. (2009) published a one-year follow-up on their participants where several score improvements were reported (Evangelini, Longley & Swartz, 2011). Eight participants completed the follow-up assessment where the amount reaching beginning proficiency for global therapist ratings, %OQ and R: Q increased by one. Those

reaching beginning proficiency for %CR increased from zero to two and remained consistent at two for %MI-Ad.

Six studies evaluated the coaching and feedback components of their training curriculums (Bohman et al., 2013; El-Mallakh et al., 2012; Forsberg et al., 2010; Robbins et al., 2011; Smith et al., 2007; Wahab et al., 2008). Throughout these feedback and coaching components, audio recordings were collected from client interviews for analysis. An additional study (van Eijk-Hustings et al., 2011) evaluated the outcomes of their training curriculum as a whole, beginning the assessment post-training and investigating changes over a six month follow-up without further guidance. The studies evaluated the MITI global ratings of MI spirit and empathy, along with %OQ, %CR, and % MI-Ad. The R: Q was reported across the research with the exclusion of van Eijk-Hustings et al. (2011).

Four studies began their assessments post-workshop and utilized the MITI to investigate MI skill throughout feedback and coaching (Bohman et al., 2013; Forsberg et al., 2010; Smith et al., 2007; Wahab et al., 2008). Results of the MITI were mixed across the research. Although increases across feedback and coaching for MI spirit and empathy were observed by Smith et al. (2007), they were not significant. In contrast Bohman et al. (2013) observed decreases on the two global ratings. Increases were reported by Forsberg et al. (2010) and Wahab et al. (2008) on the two ratings however, due to sample size, significance testing was not performed.

With regards to the summary scores, it appears Bohman et al. (2013) and Smith et al.'s (2007) participants performed the best on the %OQ and the R: Q. Smith et al. observed significant increases in these two summary scores. Increases were also detected in the research by Bohman et al., however they were not significant. The remaining summaries exhibited either increases that were not significant (Smith et al., 2007) or decreases (Bohman

et al., 2013). Increases across the summaries were again observed by Forsberg et al. (2010) however these scores were mixed in Wahab et al. (2008). For the R: Q along with %OQ and %CR, all participants increased with the exclusion of one case; having solely three participants made this a notable inconsistency. The only summary score that increased across all participants was %MI-Ad.

Two studies completed practical training components and ensured that their single participant was showing signs of proficiency prior to the feedback and coaching assessment (El-Mallakh et al., 2012; Robbins et al., 2011). In these cases MI was taught for use with a particular group of clients and the feedback and coaching was included to ensure MI skill was maintained throughout the assessments. Although already trained to proficiency, the nurses in El-Mallakh et al. (2012) and Robbins et al. (2011) continued to show signs of improvement in MI spirit and empathy throughout their assessment periods. The same was apparent for %OQ and %MI-Ad. The %CR decreased for both cases and the R: Q increased for El-Mallakh et al. but decreased for Robbins et al.

Van Eijk-Hustings et al. (2011) assessed their participants at the end of their training which had included a supervision component, and investigated changes across a six month follow-up with no additional training. Solely the mean MITI outcomes were reported in this research and no significance testing was included. Participants exhibited increases in MI spirit and empathy, however behaviour counts were mixed. Both increases and decreases were observed across the means of %MI-Ad, %OQ, and %CR.

Several studies also compared their results to the MITI proficiency thresholds. In the research by Bohman et al. (2013), proficiency was not met for any of the global and summary scores. At the final follow-up, Smith et al. (2007) reported that beginning proficiency was met solely for %OQ, although scores were close for the %CR and the MI spirit and empathy

global ratings. Beginning proficiency was met for all outcomes with the exclusion of %MI-Ad in Forsberg et al. (2010). Finally, despite the variations in scores, the nurses in the research by El-Mallakh et al. (2011) and Robbins et al. (2011) remained proficient in MI spirit and empathy throughout their assessments. Results indicated consistent outcomes for the %OQ and %MI-Ad, however %CR and the R: Q did not reach proficiency in the research by Robbins et al.

Seven studies evaluated their training curriculums using control groups, identifying effects of training through between group differences in MITI results (Daeppen et al., 2012; Forsberg et al., 2011; Moyers et al., 2008; Smith et al., 2012; Mitcheson et al., 2009; Walters et al., 2010; Young & Hagedorn, 2011). Throughout the research, recordings were utilised from client interviews, however in some cases standardised patients were used to portray clients. Several structures were used across the studies in order to evaluate the curriculums. A simple training and control group structure was implemented by Daeppen et al. (2012), Mitcheson et al. (2009), and Young & Hagedorn (2011), however Forsberg et al. (2011), Moyers et al. (2008), Smith et al. (2012) and Walters et al. (2010) included three groups in their research. Specifically, Forsberg et al. and Moyers et al. included two training groups, one received workshop training and the other received on-going training, along with a control. All the groups in Smith et al. received some form of training; one received workshop training, another received the workshop along with standard coaching (through the use of audio recordings) and the third, along with the workshop, received live coaching via teleconferencing (TCS). With regards to the research by Walters et al., along with the training group there were two controls, one group that was interested in learning MI and was waitlisted and one that was disinterested in learning MI.

The global ratings of MI spirit and empathy were reported across the research with the exclusion of Walters et al. (2010) who only reported empathy. Young & Hagedorn (2011)

separately measured the three MITI variables that comprised the spirit (evocation, collaboration, and autonomy). With regards to the summary scores %OQ, %CR, and %MI-Ad along with R:Q were reported by Daeppen et al. (2012), Forsberg et al. (2011), Mitcheson et al. (2009), Moyers et al. (2008), and Smith et al. (2012). Young and Hagedorn solely included %OQ and %CR whereas Walters et al. reported just %MI-Ad.

Of the studies utilising the simple training group and control structure, both Daeppen et al. (2012) and Young and Hagedorn (2011) reported a significant difference between their training and control group on the global MITI ratings. Mitcheson et al. (2009) observed a between groups difference on the MI spirit rating, however not on empathy. Mitcheson et al. continued to observe little difference between their groups on the summary scores whereas the trained group in Daeppen et al. performed significantly better than control on all four summary scores. Results were mixed for Young and Hagedorn, there was a significant between groups difference for %OQ, however little difference on %CR.

Results were also mixed across the studies with the more complicated structures. Forsberg et al. (2011) observed a significant difference between their on-going training group and control on the MI spirit rating; however there was little difference between workshop and on-going training. With regards to empathy, no significant between group differences were reported in this research. Moyers et al. (2008) reported similar results regarding the MI spirit rating: no significant differences between the training groups. However, in this case both training groups performed significantly better than the control group. Outcomes were consistent for empathy. Smith et al. (2012) reported results comparable with Moyers et al.; although there was no significant difference between on-going training groups on the MI spirit and empathy ratings, both groups performed significantly better than the workshop group. The trained group in the research by Walters et al. (2010) performed significantly

better on empathy than both controls, however no differences were observed across the control groups.

With regards to the summary scores, Forsberg et al. (2011) observed significant differences between on-going training and control on %CR and %MI-Ad; however there were no significant differences on the other summary scores. Moyers et al. (2008) reported no between group differences across their three groups on any summary scores and Walters et al. (2010) observed results regarding %MI-Ad that were consistent with that of empathy. Although the differences among the scores of Smith et al. (2012) for %MI-Ad were also consistent with that of MI spirit and empathy, results for the remaining three summary scores were mixed. For %OQ, the standard coaching group performed significantly better than the TCS and workshop groups. Standard coaching also outperformed TCS on %CR, however neither outperformed workshop. Finally TCS performed significantly better than both standard coaching and workshop on R: Q.

Forsberg et al. (2011), Moyers et al. (2008) and Smith et al. (2012) compared their results to the MITI proficiency thresholds. Forsberg et al. reported no groups obtaining beginning proficiency for the global or summary scores. The control group in Moyers et al. achieved proficiency on more MITI variables than the other groups: empathy and spirit, %OQ, %CR, and R: Q. The on-going training group achieved proficiency on empathy, %MI-Ad, and R: Q and the workshop group, %MI-Ad and R: Q. Smith et al. reported the percentage of participants from each group reaching proficiency on MI spirit and empathy along with %MI-Ad. Fifty per cent of TCS, 46% of standard coaching and 30% of the workshop group obtained beginning proficiency on MI spirit and empathy. Additionally, 25% of TCS and 13% of standard coaching met the level of competence for these ratings. With regards to %MI-Ad, 53% of standard coaching, 50% of TCS and 33% of the workshop group achieved the level of competence.

Across the MI training research, MITI results appeared mixed. However, the MITI has proven to be both reliable and valid (Moyers et al., 2005; Pierson et al., 2007), therefore as with the MISC, the variations among the results can be viewed as reflections of the individual training curriculums. The majority of studies that reported primarily positive results complied with the recommendations of Miller and Rollnick (2012) by having both theoretical and on-going feedback and coaching components of training (El-Mallakh et al., 2012; Forsberg et al., 2010; Robbins et al., 2011; Wahab et al., 2008; Walters et al., 2010; Young & Hagedorn, 2011) with one study spanning two and a half years (Forsberg et al., 2010). Although Daeppen et al. (2012) did not include an on-going component in their training it is likely positive results were obtained due to the assessment being carried out within one week post-workshop. MI skill can increase as a result of theoretical training, however results are not likely to remain consistent (Miller et al., 2004); if Daeppen et al. had included a follow-up period it is likely the MITI results would have decreased without coaching or feedback.

However, several studies that reported predominantly negative outcomes included coaching and feedback components in their trainings. Along with limitations unique to each curriculum such as organisational issues (e.g. Bohman et al., 2013) and small sample size (e.g. Hartzler et al., 2007), some studies reported limitations in the coaching and feedback components themselves. Forsberg et al. (2011) acknowledged the amount of time their feedback and coaching sessions were available for may not have been sufficient to effectively train their participants. Smith et al. (2007) and Smith et al. (2012) also mentioned that longer-term availability of coaching may have been required to achieve more improved overall skilfulness in MI. With regards to the research by Mitcheson et al. (2009) and Moyers et al. (2008), adherence to coaching and feedback was the concern. This training component was optional in Mitcheson et al., resulting in low attendance and submissions of in-session

recordings; making the component a requirement of training along with greater encouragement was recommended by the authors. Moyers et al. also noted a lack of motivation to engage in coaching and feedback with their on-going training participants completing only 44% of the available consult calls; difficulty accessing the coaching providers also contributed to the low engagement.

Therefore the variability in MITI outcomes adds to its validity; positive results were obtained in training programmes with sufficient feedback and coaching components. In contrast, mixed or negative results were present in curriculums that also reported limitations regarding length or adherence to coaching and feedback.

Other measures of MI skill or knowledge. Although not as well validated as the MISC and the MITI, written MI measures that assess knowledge and skill exist and are useful for evaluating the more theoretical components of training, such as workshops. The Motivational Interviewing Knowledge and Attitudes Test (MIKAT) measures MI knowledge along with beliefs consistent with the spirit of the method and the Helpful Responses Questionnaire (HRQ) along with the Video Assessment of Simulated Encounters – Revised (VASE-R) measure the ability to demonstrate MI skill through written responses. These written-based measures of MI knowledge and skill appear less frequently in the MI literature. They are useful for evaluating the increase in knowledge or the skills learnt during theoretical components of MI training however, do not provide representations of competence levels. Therefore, these types of measures appeared more commonly in the training evaluations that only involved a theoretical component such as a workshop. When a written-based measure did appear in an evaluation that included coaching and feedback, it was usually accompanying the MISC or MITI. The MIKAT, HRQ and VASE-R are described below along with psychometric information and summaries of training outcome research utilising

the various measures. The summaries are arranged the same way as the MISC and MITI: brief overviews of the studies preceded the outcome summaries.

The MIKAT. The MIKAT, developed by Leffingwell (2006), is a relatively simple test of knowledge and attitudes consistent with MI and its spirit. By administering this questionnaire before and after an MI workshop, a measure of change in attitudes and knowledge can be obtained. The MIKAT consists of two parts (see Appendix B for a copy of the MIKAT):

1. A quiz comprised of true-false questions about 10 addiction myths and four MI-consistent attitudes and assumptions.
2. A counselling behaviours checklist including five behaviours prescribed for an MI approach.

Scoring the MIKAT is straightforward; generating a summary score requires calculating the number of correct items. For more specific results, the MIKAT has been divided into the two components and a separate score for both attitudes and knowledge was interpreted. Additionally the components themselves could be divided into subcomponents therefore providing information on the amount of MI-consistent attitudes versus addition myths answered correctly in the true–false component along with the amount of correct versus incorrect behaviours selected in the checklist.

Psychometrics. Leffingwell (2006) developed and evaluated the MIKAT with child and family home-based care providers that identified encouraging results with regards to training sensitivity (described further in subsequent sections). Internal consistency of the measure has only been reported once in the literature (Doran, Hohman & Koutsenok, 2011b) revealing an adequate Cronbach's α of .84.

Literature Summary. During its development, the MIKAT was evaluated with child and family home-based care providers (Leffingwell, 2006). Participants completed a one-day (seven hour) MI workshop that included components that were both didactic and experimental. The primary focus of this training was on developing a basic understanding of MI including the spirit and strategies such as OARS. The MIKAT was completed by participants directly before and after the training. The measure identified a statistically significant improvement in correct responses in both components from pre- to post-training; a significant decrease in incorrect responses on the checklist component was also noted.

The MIKAT has only been utilised in two other known articles, both evaluating the same training programme implemented with juvenile corrections staff (Hohman, Doran & Koutsenok, 2009; Doran et al., 2011b). This workshop spanned three days and was based on the work of Miller and Rollnick (1991; 2002). The format included role-plays, demonstrations, coaching, and conceptual information. There were also orientations and examples directed at youth correctional work. More advanced two-day workshops were also included in the training, however were not described in detail and were only acknowledged in Doran et al. (2011b).

The MIKAT summary score was interpreted in both reports. Hohman et al. (2009) observed a significant increase in this summary score across training. However there was no control group included in this research, therefore there was no way to ensure that the gains made on the MIKAT were not due a practice effect. Doran et al. (2011b) investigated the effects that trainers with different levels of MI competency (members of the Motivational Interviewing Network of Trainers, MINT, versus internal trainers) may have on training outcomes. Although a significant difference was observed with MINT trainees scoring higher on the MIKAT at pre-training, results were comparable at post-training indicating that those

trained by internal trainers made greater increases on the measure. The authors also noted effects of sex and education among the scores; it appeared female trainees with more education tended to score higher on the MIKAT.

The HRQ. This measure specifically assesses the important MI skill of accurate empathy and was developed by Miller, Hedrick and Orlofsky (1991). The HRQ includes six paragraphs simulating verbalizations from individuals concerned about something in their lives. Below each paragraph a helpful response is required and the respondent is assessed on a five-point scale reflecting the extent to which reflective listening was used. Briefly, a 1 is assigned if no reflection is present in a given response but a 'roadblock' (a verbalization that interrupts the flow of conversation) is used. If a response includes both a reflection and a roadblock, a 2 is given. A score of 3 is assigned if the response is a simple reflection (i.e., effectively repeating some content of the paragraph). When a response adds inferred meaning through paraphrasing a 4 is assigned and when a response that could be scored as a 4 also includes a reflection of feeling that fits the content of the paragraph a 5 is given. The highest score is assigned when multiple responses are given.

Psychometrics. Miller et al. (1991) evaluated the reliability and internal consistency of the HRQ. Inter-rater reliability was computed using product moment correlations across two independent raters. The reliability coefficients for each item were excellent (.71–.91) as too was the reliability for the measure as a whole, at .932. Test-retest reliability was also evaluated by correlating HRQ scores both pre- and post-training in MI. An adequate correlation coefficient of .45 was obtained from the 120 individuals being evaluated. Internal consistency of the HRQ was also adequate. Cronbach's α coefficients for pre- and post-training were .92 and .89 respectively.

Adaptations of the HRQ. The simulated interactions comprising the HRQ reflect predominantly substance abuse scenarios that most commonly would be encountered by a drug and alcohol counsellor. With the use of MI within the criminal justice system increasing, it was decided that a measure specifically designed to rate probation interactions was required. Therefore the Officer Responses Questionnaire (ORQ) was derived from the HRQ as a reflective listening measure especially for the work of probation officers (Walters, Alexander & Vader, 2008). The ORQ is of the same format as the HRQ, with five scenarios that require a written response. These scenarios reflect common situations in probation such as substance abuse, aggression, and obtaining employment. The five-point scale was maintained from the HRQ; however modifications were made to the scoring guidelines. The primary alterations included the categorization of reflections into simple and complex along with no penalization for question asking. The scoring guidelines are as follows (Walters et al., 2008): a 1 is assigned if a roadblock is included, a 2 if the response is not reflective, however includes verbalizations such as affirmation or a closed question. If an open-ended question is included a score of 3 is given, a simple reflection receives a 4, and a complex reflection, a 5. Again, if multiple responses are given the highest possible score is marked, however if any are roadblocks, the answer is spoiled and a 1 must be assigned. Inter-rater reliability for the ORQ was computed using ICCs and revealed good to excellent results (.73–1.00, Walters et al., 2008).

The ORQ was itself adapted for juvenile corrections, thus forming the Workers Responses Questionnaire (WRQ, Hohman et al., 2009). The scenarios were modified to reflect interactions with juveniles consistent with what a probation officer may experience. No explanation of the WRQ scoring guidelines exist, therefore it was expected they were consistent with the ORQ. Inter-rater reliability was computed using ICCs (Doran et al.,

2011b) and ranged from fair to excellent (.45–.91). Internal consistency was also computed by these authors, revealing an adequate Cronbach's α coefficient of .82.

Literature Summary. The HRQ was developed much earlier than other written measures of MI skill and has thus been utilised more frequently throughout the literature. Six studies used the HRQ in evaluating their training curriculums (Baer et al., 2004; Childers et al., 2012; Cucciare et al., 2012; Martino, Haeseler, Belitsky, Pantalon & Fortin IV, 2007; Miller et al., 2004; Young & Hagedorn, 2012). Across these studies a single summary score produced by the measure has been most commonly used (average score of the six vignettes), however results have also been interpreted in further detail in order to capture all the variables measured by the HRQ.

The HRQ has been used as a secondary measure alongside the MISC and MITI (Baer et al., 2004; Miller et al., 2004; Young & Hagedorn, 2012) and in one case an unnamed coding system drawn from and guided by the MITI (Cucciare et al., 2012). In all studies a significant increase from pre- to post-training was observed, however Young & Hagedorn (2012) reported an additional surprising outcome. Although their training group significantly improved on the HRQ, so did their control resulting in little post-training difference between groups. This result was inconsistent with that of the MITI, where most variables exhibited significant between group differences (described in earlier sections). The authors noted that this outcome may have been because the HRQ measures more variables that were not individually investigated; by being more specific in the interpretation of the measure, results may have been different. This pattern of results could also be an indication of a practice effect; perhaps the gains made on the HRQ from both groups were a reflection of having already had experience with the measure.

When utilised as a primary measure, each variable on the HRQ was measured separately for more detailed interpretation (Childers et al., 2012; Martino et al., 2007). These studies measured the changes in participants open and closed questions, reflections, and roadblocks. Childers et al. (2012) also interpreted MI-consistent strategies whereas Martino et al. (2007) included a measure of reflection depth, obtained by summing scores across vignettes. Childers et al. observed significant increases across open-ended questions, reflections, and MI-consistent strategies along with significant decreases in closed-questions and roadblocks. Although all score movements were in the desirable directions in Martino et al., the increase in open-ended questions was not significant.

Adaptations of the HRQ have been used more sparingly throughout the literature, The ORQ has been utilised twice (Walters et al., 2008; Walters et al., 2010) and the WRQ has also had two publications. However, both articles that used the WRQ were evaluating the same training programme for juvenile corrections staff (Hohman et al., 2009; Doran et al., 2011b), the research that also made use of the MIKAT.

As part of its development, the ORQ was administered before and after a two-day MI training programme carried out across several American federal districts (Walters et al., 2008). Significant increases were observed across all five items on the measure. The ORQ was also used alongside the MITI when evaluating an MI training programme for probation officers (Walters et al., 2010). Results were consistent with the two MITI variables interpreted in the study (empathy and MI-Ad); there was a significant increase from pre- to post-training that resulted in a significant between groups difference when compared to control.

With regards to the WRQ, a significant increase was reported by Hohman et al. (2009) as a result of their training programme. When comparing the effects of MINT versus

internal trainers, little difference was observed on the WRQ (Doran et al., 2011b). However, it was noted that MINT trainees had on average higher scores at both pre- and post-training. The authors also identified effects of sex and education among the scores; as with the MIKAT, female trainees with higher education tended to have higher scores on the WRQ.

The VASE-R. The VASE-R, a video-based assessment of MI skills, provides a standardised MI challenge and assesses a variety of MI skills (Rosengren, Hartzler, Baer, Wells & Dunn, 2008). This measure consists of three video-based scenarios with different substance abusing clients played by actors. After each client there are questions that require written responses which are scored against MI standards. There are 18 questions that cover skills within five subscales: reflective listening (RL), responding to resistance (RR), summarising (S), eliciting change talk (ECT), and developing discrepancy (DD; see Appendix C for VASE-R questions and sample answers).

Of the 18 items, 15 are of a free-response format with the remaining three being multiple choice (Rosengren et al., 2008). Items are scored using a 3-point system; what constitutes a certain amount of points varies across subscales. In regards to RL, RR, and ECT, a 0 is given if the response is likely to evoke resistance or is confrontational; if the response is an inaccurate representation of the client's speech or neutral, a 1 is given; if there is an accurate representation of the intended MI skill and reflection of the client's speech, a 2 is given. For S, a 0 is given if the response is again confrontational or likely to evoke resistance, but also if it does not include multiple ideas. A 1 is given if the response has multiple ideas, however only includes either ambivalence or client change talk. It is only when the response includes multiple ideas along with both ambivalence and client change talk that a 2 is given. Finally for the DD subscale, a 0 is given if an incorrect option is chosen and paired with a rationale not consistent with MI; a 1 given if again an incorrect option is

chosen, however the rationale is MI consistent; for 2 points the correct option must be chosen. The collated points yield a full-scale score that can range from 0–36 (see Appendix D for further scoring guidelines).

A New Zealand edition of the VASE-R has been developed using actors with New Zealand accents to portray clients (Hall McMaster & Associates Limited, 2012). The content of the DVD is analogous to that of the original measure and the original scoring form, manual and answer sheet are used (Rosengren et al., 2005; 2009).

Psychometrics. Through two large scale studies of the VASE-R, the psychometrics were obtained along with proficiency standards (Baer et al., 2008, as cited in Rosengren et al., 2008; Rosengren et al., 2008). From research with substance abuse treatment personnel participating in a trial comparing different MI training methods (Baer et al., 2008, as cited in Rosengren et al., 2008), the inter-rater reliability, and internal consistency along with concurrent validity were established. Descriptive statistics for the VASE-R were completed to assist in the development of proficiency standards (described further in subsequent sections). The research participants completed the VASE-R, along with the HRQ and interviews with standardised patients that were coded by the MITI; these additional methods are comparable to the VASE-R as they are established measures of MI skilfulness.

Inter-rater reliability was evaluated through the computation of ICCs for each subscale. All ICCs varied across subscales, however were all greater than .40 and therefore deemed acceptable when compared to published guidelines (Cicchetti, 1994). Internal consistency, assessed using Cronbach's α coefficients, was acceptable for the total scale ($\alpha = .85$) however, varied across subscales: the ECT and DD subscales in particular were $\alpha = .49$ and $\alpha = .44$ respectively. As the acceptable minimum alpha for psychological measures in newer areas of research is .70 (Lance, Butts & Michels, 2006), these are extremely low.

Concurrent validity was conducted by correlating participant scores from the VASE-R with that of the HRQ and MITI. Full-scale and subscale VASE-R scores positively correlated with the single summary score produced by the HRQ along with various global and behavioural MITI scores.

Norms and proficiency standards for the VASE-R were established from administering the measure at an annual meeting of the Motivational Interviewing Network of Trainers (MINT, Rosengren et al., 2008), along with the descriptive statistics from the research of Baer et al. (2008, as cited in Rosengren et al., 2008). Members of MINT are considered skilful and knowledgeable in MI, therefore were ideal for the creation of proficiency standards.

Descriptive statistics for this participant sample were completed and compared to that of the previously described research (Baer et al., 2008, as cited in Rosengren et al., 2008). The mean full-scale score from the MINT participants was higher than the substance abuse treatment practitioners at baseline (30.88 versus 18.09). The MINT members' mean score also exceeded the post-training score of the substance abuse practitioners (24.13). These sample differences were consistent across the subscales and vignettes.

Rosengren et al. (2008) used the scores generated by the two samples to determine reference points for the VASE-R. Three benchmarks were developed: untrained, beginning proficiency and expert proficiency. The MINT members represented "expert proficiency" and the substance abuse treatment practitioners prior to training represented "untrained". The authors then used statistical methods developed by Jacobson et al. (1999, as cited in Rosengren et al., 2008) and Jacobson and Truax (1991, as cited in Rosengren et al., 2008) to define "beginning proficiency" by identifying points between the untrained and expert proficiency distributions. From this procedure the benchmarks and proficiency standards

were developed: untrained (full-score = 18), beginning proficiency (full-score = 26) and expert proficiency (full-score = 31).

Literature Summary. As part of the development of the VASE-R (Rosengren et al., 2008) sensitivity to training effects were assessed. Substance abuse practitioners completed the measure before and after MI training and at a three month follow-up. The 15-hour training compared two models that differed with regards to format (two days versus five 2-3 hour sessions) and the use of “medical actors” for skills practice. Further details of these trainings were outlined in Baer et al. (2009). Despite the differences in training, Rosengren et al. (2008) stated that both methods lead to comparable results. The trainings did not differ with regards to content and covered MI concepts such as the spirit, OARS and MI-consistent strategies. There were also practice opportunities where trainers provided feedback.

Both the total VASE-R score and the individual subscales were included in the analyses. There was a significant increase in the VASE-R total score from pre- to post-training. A small decline, which was not significant, was identified from post-training to follow-up. With regards to the individual subscales, all exhibited statistically significant increases from pre- to post-training. No change was identified from post-training across follow-up with the exception of S, which significantly decreased.

Two further training curriculums have utilised the VASE-R. Whilst training medical students in MI, the first vignette of the VASE-R was used to evaluate progress (Bell & Cole, 2008). The training itself included four weekly two-hour sessions and comprised of didactics, role plays, video demonstrations and interactive exercises. Core MI principles were adapted to suit general medical practice through the addition of mutual agenda setting and individualised health related feedback components.

Across training a significant increase in the total vignette score was observed. The authors also investigated the changes in each subscale; significant increases were identified for all subscales with the exception of DD. Although not mentioned in the study, the DD subscale is measured as multiple choice which could have explained the lack in increase; simply checking the wrong answer results in a zero.

Doran, Hohman and Koutsenok (2011a) solely measured the total VASE-R scores in their evaluation of the training for juvenile correction staff. This training has been detailed in earlier sections as it was also analysed by Hohman et al. (2009) and Doran et al. (2011b). A significant increase as a result of training was identified. Additionally, demographic variances among the participants were investigated with regards to VASE-R scores; although not associated with the pre- to post-training increase, effects of both age and education were observed. Specifically, younger and more highly educated participants achieved higher VASE-R scores at both pre- and post-training.

Overall, the VASE-R appears to effectively capture the effects of MI training, however again the possibility of the practice effect cannot be ignored. No research utilising the measure included a control group therefore it cannot be ensured that the gains made were a result of just the training programmes. Further research is required that controls for the possibility of this practice effect.

Although not commonly utilized throughout the literature, written measures of MI knowledge or skill have been observed to effectively evaluate theory-based trainings. All studies that used written measures with the exception of Rosengren et al. (2008), made their evaluations directly post-training. Therefore, the predominantly positive results were consistent with reports by Miller and Rollnick (2012) that initial gains in MI knowledge or skill are possible as a direct result of workshop training. The decrease in VASE-R skill across

follow-up observed by Rosengren et al. was also consistent with the common observation of skills when no feedback or coaching is provided as part of training. The measures proved useful in making initial assessments regarding MI knowledge or skill in ways more efficient than the MISC or MITI and can be utilised preceding or alongside these measures to achieve a more complete assessment of training progress.

There are few established measures of MI knowledge, skill, adherence, and performance. Of this small selection, the most reliable and validated are also the most labour intensive. Although the simpler, more efficiently administered MI measures have initially positive supporting research, further investigations are required; there is a lack of the necessary revision and refinement across these assessments. The present research is an evaluation of two of the lesser researched MI written measures: the MIKAT and VASE-R.

The occurrence of an MI training programme for staff of Child Youth and Family services (CYFs) New Zealand provided outcome data of the MIKAT and the VASE-R for the present analysis. MI was identified as a key skill useful for implementation with youth residing at the CYFs Youth Justice and Care Residences. These residences are safe secure places where youth who have been placed in CYFs care or after committing serious offences can get support to change their lives. Within the residences, help is provided for behaviours that could have contributed to offending such as substance abuse or problems with anger. The youth must attend school every day; there is a school building and open areas for outdoor activities within the facility. Plans are also formulated for when the young person leaves the residence, whether it is to get a job, go back to school, or participate in treatment mandated by the Youth Court (Cyf.govt.nz, 2013).

The desired goals of implementing MI at the residences include the improvement of behaviour whilst in CYFs care, along with improving engagement in services and

programmes after release potentially leading to the reduction in the risk of reoffending. A training and implementation programme was developed and comprised of a two-day MI workshop completed by all CYFs staff working with youth at the residences. More senior staff continued on to complete a one-day advanced workshop, followed by feedback and coaching based on audios of MI sessions with youth at the residences.

The primary goal of the present research was to evaluate the MIKAT and VASE-R using data from the CYFs MI training programme (described in earlier sections). Because the measures are under researched with no investigations carried out in New Zealand, full psychometric evaluations were conducted to determine whether the MIKAT and VASE-R were in fact reliable and valid. These measures were selected because they were rarely utilised across the MI literature, therefore an investigation into their psychometric properties was warranted. Additionally, each measure has unique elements making them potentially desirable measures for future training evaluations; the MIKAT is the only known officially developed measure of MI knowledge and attitudes and the VASE-R aims to approximate client interactions through video simulations. The unique properties of the MIKAT and VASE-R further emphasized the importance of investigating their reliability and validity, and determining if they would be effective tools for future training evaluations. If the measures are identified to be reliable and valid and their use across the literature is increased, evaluators will have more measures to select from when developing a training and implementation plan. Increased use of these written-based measures may make evaluations of MI training less labour-intensive.

Method

As stated previously, Miller and Rollnick (2012) suggest guidelines for effective MI training: initial training to increase MI knowledge can be in the form of a workshop or self-study, however skill cannot be taught under these circumstances. Therefore, coaching and feed-back need to also be provided to ensure skill attainment. The MI training evaluated by the present research in order to investigate the MIKAT and VASE-R was designed in accordance with these guidelines and included two workshops along with on-going coaching and feed-back.

Procedure

Training workshops were completed between May 2012 and July 2013. CYFs staff from five regions across New Zealand attended the two-day MI training. The member of MINT travelled to each location to provide the workshops. Because the most recent revision of the MI spirit and structure was published on August 14th, 2012 (Miller & Rollnick, 2012) four of the five initial workshops were taught to the prior guidelines (Miller & Rollnick, 2002). The fifth workshop was initiated much later than the others because it was carried out specifically to train newly employed staff, all in the Auckland region. Prior to all initial workshops, participants were invited to read two MI articles for the purpose of gaining background knowledge: Ten Things Motivational Interviewing is Not (Miller & Rollnick, 2009) and Towards a Theory of Motivational Interviewing (Miller & Rose, 2009). Participants completed the MIKAT immediately prior to and after the workshop.

More senior CYFs staff participated in the one-day advanced training. The amount of time between initial and advanced trainings varied for each location. Christchurch had seven months between workshops, Wellington, twelve months, and Hamilton, thirteen months. With regards to Auckland, the original group who received the first initial workshop had

eleven months before the advanced training. The group who received the second workshop had just a week between trainings. The same MI trainer again travelled to each location for training administration. Participants were updated on the revised spirit and structure of MI (Miller & Rollnick, 2012) during this workshop. Prior to the one-day workshop, participants received a copy of the MITI (Moyers et al., 2010) so they would be familiar with the measure assessing them in practice. Participants completed the VASE-R both before and after this workshop.

The time between the advanced MI workshop and the beginning of the feedback and coaching varied depending on when each senior staff member submitted their first audio. After each audio was coded, written feedback was developed based on the MITI summary scores and individual coaching by the coder was provided via telephone.

Data from the MIKAT, VASE-R, and MITI were collected for analysis. With regards to the VASE-R, raters must reach a certain level of reliability as the measure is primarily free response. The VASE-R Administration and Scoring Manual (Rosengren et al., 2005) stated that 90% inter-rater agreement is recommended. The rater had no experience with the VASE-R prior to the present research, therefore training was provided by a member of MINT in order to reach this level of agreement. The rater and the member of MINT reviewed the general concepts of MI along with the various skills required to effectively perform the method. The VASE-R was then completed by the rater and the responses were discussed with the member of MINT whilst reviewing the VASE-R manual. The VASE-R was also completed by a third party for the rater and the member of MINT to score and therefore assess their level of agreement. Once the desired level of agreement was achieved, the rater was permitted to begin collecting data from the participant VASE-Rs.

A second member of MINT evaluated the in-session recordings with the MITI and the data were then passed on for analysis in the present research. The implementation plan for the training included the provision of feedback and coaching, therefore the member of MINT was required to carry out the MITI coding.

Participants

Eighty-five members of CYFs staff from the five regions of New Zealand completed the initial two-day MI workshop. All staff employed to work with the youth at the residencies were invited to attend training. The senior members of staff continued on to complete the one-day advanced workshop ($N = 46$) and receive feedback and coaching from in-session audios (see Table 1 for demographic details). It proved difficult to retrieve demographic information beyond the gender of the participants. Out of 46, ages could only be found for 36 participants. Ethnicities were retrieved for 35, education for 24 and job titles for 26. The demographics are reported based on the above numbers of participants. Participant's mean age was 42.17 and 30.43% were male. In terms of ethnicity the majority (40%) identified as New Zealand European. A similar amount of participants identified as either New Zealand Māori or Pacific Islander and a small amount identified as being from Africa or the United Kingdom. Reports on education identified that most of the participants had either undergraduate or postgraduate university qualifications. There were, however three accounts of education that was either of "college", "university entrance" or solely stated the name of the university they attended. It was not obvious exactly to what level of education these individuals had received; therefore they were marked as being unclear. Finally with regards to job title, exactly half of the participants that reported a job title referred to themselves as a Case Leader. The other half of these participants identified themselves as being a Social Worker or Youth Worker, a Team Leader of Clinical Practice or a Senior Practitioner.

Table 1. Participant demographic characteristics

Characteristic	Participants who attended advanced training % (actual)
Age ^a	
M	42.17
SD	10.84
Range	24 – 65
Gender	
% Male	30.43 (14)
Ethnicity ^{bc}	
New Zealand European	40.00 (14)
New Zealand Māori	22.86 (8)
Pacific Islander	20.00 (7)
African	8.57 (3)
British	8.57 (3)
Education ^d	
Undergraduate Degree or Diploma	58.33 (14)
Postgraduate Degree or Diploma	29.17 (7)
Unclear	12.50 (3)
Job Title ^e	
Case Leader	50 (13)
Social Worker	15.38 (4)
Senior Practitioner	3.85 (1)
Team Leader of Clinical Practice	15.38 (4)
Youth Worker	15.38 (4)

^aN=36 ^bIf a participant identified with more than one ethnicity, the first was recorded in the above table. ^cN=35. ^dN=24. ^eN=26.

Implementation Plan

The initial workshop. The initial MI training was a two-day (14 hours) workshop conducted by the member of MINT. The training consisted of video-taped demonstrations, modelling, didactic teaching along with real-play and role-playing with feedback, consistent with the recommendations of Miller and Rollnick (2002). The content of this workshop focused on the MI spirit and the skills of OARS along with emphasis on the concepts of resistance, ambivalence, and change talk (refer to the spirit and structure of Motivational Interviewing sections of the Introduction for further details). The principles of MI (express empathy, support self-efficacy, developing discrepancy, roll with resistance, and avoid argumentation) were also included along with research presented to demonstrate the effectiveness of the method and how it works. There was also an introduction to MI strategies

such as the importance of confidence rulers. The workshop that was administered after the release of Miller and Rollnick's (2012) most recent MI text followed the same structure and included much the same content as the prior workshops. However, newer resources and revisions such as the extension of the spirit along with the latest MI structure (engaging, focussing, evoking, and planning) were included to make the content consistent with Miller and Rollnick (2012).

The advanced workshop. The one-day (7 hours) advanced MI training was conducted by the same member of MINT. This training again consisted of video-taped demonstrations, modelling, didactic teaching, and role-playing with feedback, consistent with Miller and Rollnick's (2012) recommendations. This workshop included a re-cap on the concepts of ambivalence and change talk and the importance of the role of these concepts in MI. There was also a reminder of the spirit and its importance. The focus of this training was around the practice of MI skills, in particular reflective listening as a means of generating, evoking, and responding to change talk.

Coaching and feedback. Participants were invited to produce in-session recordings as demonstrations of their use of MI with the youth in the residencies. Recordings were sent to the second member of MINT for evaluation. Each participant then received feedback and coaching based on the recording. The intention was that the process would be repeated four times to evaluate participant progress in practicing MI.

Measures

Assessment of MI knowledge. Miller and Rollnick (2012) advised that it is important to teach about MI and its spirit in a given training, regardless of context. As mentioned previously, it was decided to analyse the MIKAT as it lacked in supportive research and psychometrics, having only been used in published studies twice after its development. Also,

in these publications the measure was successfully implemented with juvenile correctional staff (Hohman et al., 2009; Doran et al., 2011b), a similar participant sample to the present training, providing indications for its effectiveness with criminal justice populations. The measure is also simple and relatively fast to complete (approximately five minutes); if identified as reliable the MIKAT would be an ideal measure for quick assessments of introductory MI workshops.

The true–false component of the MIKAT is comprised of statements identifying clients as substance abusers. Therefore, as the clients of the CYFs staff were youth who may or may not have had a problem with substance use, modifications were made to the vocabulary of the measure to better portray clients in general (see Appendix E for the MIKAT modified for use with CYFs).

The initial two-day workshop outlined in the implementation plan had a primary focus on educating about the spirit and structure of MI. This focus made the training component appropriate for evaluating the validity of the MIKAT.

Evaluating MI skill. Once knowledge of MI and its spirit have been established, Miller and Rollnick (2012) then advised the development of MI skills. As mentioned previously, the most reliable and valid measures of MI skill are those that make the assessment through actively observing MI in practice. These measures are labour intensive and require the recording of MI sessions. Written measures that can assess skill prior to any practical use of MI could be more efficient for evaluating training curriculums. Of the few written measures evaluating MI skill, the VASE-R was selected for use. Although psychometric evaluations have been previously carried out (Rosengren et al., 2008) the VASE-R has had few publications. The measure is also unique because it has a more modernised design to simulate client interactions through the video presentation of actors

portrayed as clients. Having to listen to the client and abide to specific time limits for each question, the measure better approximates actual MI sessions. The New Zealand edition of the VASE-R (VASE-R NZ) was implemented in the present research, making the evaluation the first for this edition. The focus of the advanced MI workshop was on the development of MI skill. Therefore this component of training was appropriate for evaluating the validity of the VASE-R.

Measure of MI performance. The final component of training recommended by Miller and Rollnick (2012) is feedback and coaching based on observations of MI performance to fully develop skill and adherence. The MITI was selected as part of the training curriculum to evaluate the in-session MI recordings and coach the participants through their final component of training. The MITI was chosen over the MISC for several reasons. Being derived from the MISC, the MITI can be viewed as an improvement, having established stronger reliability along with being less labour intensive (Moyers et al., 2005). The MITI is also more condensed than the MISC, yet continues to cover all the necessary aspects of MI such as the spirit and unique practitioner behaviours. Being used as part of a training curriculum meant only the practitioner variables were of interest; the MITI provided a means of assessing these variables without unnecessarily measuring client responses.

As the MITI was well validated, no further evaluation was included in the present research. However, it was planned to utilise the data from the participant's first audios to further evaluate the validity of the VASE-R.

Data Analysis

The aim of the present research is to determine whether the MIKAT and VASE-R are reliable and valid measures of MI knowledge and skill. Therefore, using the scores from the training participants, full psychometric evaluations of the measures were carried out through

the following set of analyses. As part of this investigation, sensitivity to the effects of training was also assessed.

The MIKAT. Psychometric evaluation of the MIKAT included examining descriptive statistics for the individual items, components, and total scores were computed for both pre- and post-training. Analysis of the descriptive statistics allowed for identification of variation in difficulty among the items and components.

Internal consistency of the total MIKAT and individual components for both pre- and post-training was assessed using Cronbach's α coefficients and scale correlations. The suggested minimum acceptable α value for psychological measures considered new is .70 (Lance et al., 2006).

The MIKAT has had two publications since the one detailing its development (Leffingwell, 2006). After development, Leffingwell (2006) evaluated the measure, dividing it up into the two components and investigating both changes in correct responses and incorrect responses. The additional two publications detailed different aspects of the same study (Hohman et al., 2009; Doran et al., 2011b) and used only the total MIKAT score to assess their participants. It was thought necessary to investigate whether it was appropriate to analyse the measure as a single score or to divide it up, therefore making it possible to observe changes in both correct and incorrect responses.

In addition to Cronbach's α coefficients, a principal components analysis (PCA) was conducted. A PCA is a technique of reducing variables into smaller sets or factors that represent the majority of the information from an original set by examining the structure of the relationships among the variables (Norman & Streiner, 1998). In the present analysis a PCA determined how many components were explained by the MIKAT through component loadings and observing the amount of variance explained by each component. It is sometimes

common for a parallel analysis (PA) to precede a PCA, however due to the small number of variables making up the MIKAT, it was not appropriate to perform a PA.

After identification of the number of components for analysis of the MIKAT, Mann-Whitney U tests were performed to determine whether there was a significant difference between pre- and post-training scores. Confidentiality restricted the pairing of scores and the ownership of missing tests, therefore a nonparametric test was deemed appropriate. A power analysis determined that for 95% power, a sample size of $N = 110$ at both pre- and post-training was required assuming a medium effect size ($d = .50$). However, power was significantly reduced due to missing data.

Prior to the analysis, a Little's Missing Completely at Random test (MCAR; Little, 1988) was run to determine whether the data were missing at random or if there was a systematic pattern among the incomplete cases. The MCAR test identified that the data were not missing completely at random; there was a pattern amongst the missing data. The pattern emerged because of the simplicity of the data, solely measuring pre- and post-training. A pattern was visible because several participants either did not complete the pre-training or post-training assessment.

However, as an exploratory analysis, multiple imputations was used to replace missing values in the pre- and post-training datasets to identify whether the results changed and thus whether any concerns regarding the missing data could be identified. This technique involves generating imputations for missing data several times, each time with different imputations and then combining the estimates to result in a complete dataset (Fichman & Cummings, 2003). There was expected to be little difference between the two tests as the individuals with missing data were not believed to differ from the rest of the sample. The

Mann-Whitney U test was run twice, once with the incomplete cases excluded and again using the multiple imputations dataset.

Finally, a Pearson correlation was performed using the total MIKAT scores to determine the relationship between the pre- and post-training results. Agreement between the pre- and post-total scores was then measured using the Gower coefficient of agreement (Gower, 1971). Briefly, the Gower coefficient index expresses the average absolute discrepancy between pairs of observations (scaled relative to the maximum possible discrepancy), which is then re-explained as a measure of agreement by subtracting the discrepancy value from one. It varies between zero and one, with one indicating identity between the observations. As the initial workshops were implemented across several locations over an extended period of time (approximately 12 months), high agreement on the Gower coefficient would indicate that all groups improved similarly on the MIKAT regardless of location or time of workshop.

The Gower index was computed using the “Gower” computer software version 1.1 (www.pbarrett.net/software.html). The coefficient has been used recently in research comparing family and self-ratings of injury awareness in individuals with traumatic brain injury (Barrett, McLellan & McKinlay, 2013). Because there is no obvious hypothetical sampling distribution for the Gower coefficient, a bootstrap procedure was employed to compute credibility intervals (the interval where it is expected to observe 95% of all coefficients computed using the same sample size, number-type and same minimum and maximum possible data range as that for the observed coefficient). A total of 10,000 resamples, of the same sample size used in the present study, from a uniform random number distribution were undertaken, from which the empirical sampling distribution of coefficients was created and against which the observed coefficient could be assigned a probability of

occurrence (the significance test) and an appropriate credibility interval constructed. Full details of the exact procedure are contained in 'Bootstrap' version 1.0, the programme used to perform the procedure (www.pbarrett.net/Bootstrap/Bootstrap.html).

The VASE-R. Psychometric evaluation of the VASE-R consisted of a similar sequence of computations to that of the MIKAT. Two-way mixed, absolute agreement, average measures intra-class correlations (ICCs) were used on the subset of the measures to assess the degree to which the raters demonstrated agreement in their overall and subscale ratings of the VASE-R. ICCs describe relationships between different groups. With regards to measuring inter-rater reliability, ICCs demonstrate the level to which independently rated groups of scores are related; stronger relationships result in larger ICC values.

Descriptive statistics for the individual items, subscales, and total scores of the VASE-R were computed for both pre- and post-training. Analysis of the descriptive statistics allowed for identification of variation in difficulty among the items or subscales.

Internal consistency of the total VASE-R and the subscales for both pre- and post-training was evaluated using Cronbach's α coefficients and scale correlations. When Cronbach's α coefficients are low any scale correlation is attenuated. Specifically, the magnitude of the correlations between variables reduces due to measurement error (Muchinsky, 1996). Thus, the validity estimate is reduced, creating bias. The formula for disattenuation provides an estimate of the correlation if the scores for both measures had perfect reliability. Accordingly scale correlations were disattenuated using the 'Correlation Attenuation Correction' software version 3.0 (www.pbarrett.net/software/atten3.zip).

In researching the reliability and use of the VASE-R, uncertainty arose about whether the VASE-R should be analysed using the total scores or divided into five subscales. Specifically there was inconsistency across the few studies that had used the measure with

one reporting the total scores (Doran, Hohman & Koutsenok, 2011a) and the other reporting individual subscale results along with the total scores with no justification as to why this was done (Bell & Cole, 2008). Thus it was also decided to investigate the most appropriate analysis of the VASE-R.

In addition to Cronbach's α coefficients, a parallel analysis (PA) was conducted. A PA is used to determine the number of factors to retain through the generation of random variables based on Monte-Carlo simulations (Ledesma & Valero-Mora, 2007). Here eigenvalues are generated, which if statistically significant, indicate the number of components to retain; only factors with eigenvalues that are unlikely due to chance are desired. Eigenvalues extracted from a correlation matrix of an analysed measure are compared with those from normal values that are uncorrelated. A PA offers a 95% confidence interval for five eigenvalues and only those factors that exceeded this confidence interval were retained for use in the PCA.

Commonly used statistical software packages do not include the option for PA (e.g. SPSS) however they do permit syntax programming. Thus syntax created by O'Connor (2000) was edited and implemented in SPSS to enable the use of PA (<https://people.ok.ubc.ca/briocconn/nfactors/nfactors.html>).

Several specifications are made using syntax including the number of parallel datasets for analysis: the percentile of the distribution and random data eigenvalues, whether the analysis is run as a precursor to a PCA or a common factor analysis, and whether normally distributed random data generation or whether permutations of the raw data is used. For the present analysis 1000 parallel datasets were used and the 95th percentile was chosen for the distribution and random data eigenvalues (equivalent to $p=.05$). The PA was the precursor to a PCA and, as the data were normally distributed, the random data generation option was

chosen. In running the PA, one component was identified as significant and therefore retained as the number of factors to extract in the PCA. The PCA was then performed to determine the amount of variance explained by the one-component model.

After identification of the correct method of analysis for the VASE-R, the computation of concurrent validity was planned to demonstrate how well the measure correlated with a well validated evaluation of MI skill and adherence. The VASE-R total scores were to be correlated with the global ratings of the MITI (see Introduction for MITI description). Pearson correlations were to be used for the analysis. A power analysis determined that for 95% power, a sample size of $N = 115$ was required in order to observe a correlation of at least .30, the average correlation observed by Rosengren et al. (2008) when computing concurrent validity with the same measures. However, the amount of participants that submitted post-training VASE-Rs was little over a third of this recommendation ($N = 38$), and because only six of these participants submitted audios for MITI evaluation in time for the present research, it was decided to omit this analysis.

To determine whether the VASE-R was sensitive to training effects, a paired samples t-test was performed to identify whether there was a significant difference in scores between pre- and post-training. A power analysis determined that for 95% power, a sample size of $N = 54$ was required assuming a medium effect size ($d = .50$). However, the present sample was smaller than required ($N=38$) and there were also data missing which significantly reduced power.

An MCAR test (Little, 1988) identified that the data were not missing completely at random and there was a pattern amongst the missing data which was consistent with that of the MIKAT. Multiple imputations was used as an exploratory analysis to replace missing values in the pre- and post-training datasets to identify whether the results changed and thus

whether any concerns around the missing data could be identified. Again, little difference between the two t-tests was expected as the individuals with missing data were not believed to differ from the rest of the sample. The t-test was run twice, once with the incomplete cases excluded and again using the multiple imputations data set.

As a further exploratory analysis, a multiple regression was planned to identify whether high scores on the VASE-R predicted high scores on the MITI. This was to be done using the post-training VASE-R scores and the first intake of MITI results. A power analysis determined that for 95% power, a sample size of $N = 89$ was required assuming a medium effect size ($f^2 = 0.15$). However, again the amount of participants submitting post-training VASE-Rs was under half of this recommendation ($N = 38$), and because only six of these participants submitted audios for MITI evaluation in time for the present research, it was decided to omit this analysis.

Finally, Pearson correlation was performed, this time to determine the relationship between the pre- and post-training results. The correlation was then disattenuated to account for measurement error (as previously outlined in earlier sections) and agreement between the pre- and post-total scores was measured using the Gower coefficient of agreement (Gower, 1971) and the bootstrapping procedure to obtain significance. The advanced workshops were implemented across the four locations, this time over seven months. Again, a high Gower coefficient would indicate similar improvement across all groups regardless of location of the workshops.

All data, with the exclusion of the disattenuation and Gower calculations, were analysed using SPSS 20 for windows (IBM Corporation, Armonk, NY, USA). An alpha level of .05 was used for all statistical tests.

Results

The MIKAT

Descriptive statistics. Table 2 lists the means and standard deviations for the individual items, two components, and total MIKAT scores for pre- and post-training. Results are displayed with regards to the mean scores of each component, subcomponent and subsequent items. With regards to the individual items, the mean scores represent the percentage of participants who answered the items correct, for example the increase across time points for item one in the true–false component indicates that there was an increase in participants correctly marking these statements as false (e.g., 51.4% - 74.4%) thus contributing to their overall score. This is with the exception of the incorrect responses in the checklist component which demonstrates a decrease from pre- to post-training. This decrease reflects a reduction in the selection of incorrect responses.

Results reflect potentially higher levels of difficulty among individual items. For example, correct responses for item 12 (“Readiness to make change is the clients responsibility – no one can help them until they decide they are ready”) were consistently low across time points ($.351 \pm .481$ – $.346 \pm .479$). Results appeared similar for item 21 in the checklist component (“Develop Discrepancies”); although there was an increase from pre- to post-training ($.270 \pm .447$ – $.667 \pm .474$), results for this item remained low in comparison to the four other correct options.

What is also evident in Table 2 is the variability across the results; it appears there were no items that all participants answered correct. However, items five and six (Substance users need to hit “rock bottom” before they can change; If clients are resistant to talk about changing their behaviour, direct confrontation and persuasion are required to help the person

Table 2. Means and Standard Deviations for Items, Components, and Total MIKAT scores for Pre- and Post-Training

	Pre- (N=74)	Post- (N=78)
	M \pm SD	M \pm SD
True-False	8.135 \pm 3.036	10.231 \pm 2.301
Myths	5.838 \pm 2.938	7.487 \pm 2.005
Item1	0.514 \pm 0.503	0.744 \pm 0.439
Item2	0.257 \pm 0.440	0.551 \pm 0.501
Item3	0.662 \pm 0.476	0.603 \pm 0.493
Item5	0.811 \pm 0.394	0.923 \pm 0.268
Item6	0.770 \pm 0.424	0.923 \pm 0.268
Item7	0.392 \pm 0.492	0.667 \pm 0.474
Item9	0.527 \pm 0.503	0.833 \pm 0.375
Item11	0.730 \pm 0.447	0.897 \pm 0.305
Item12	0.351 \pm 0.481	0.346 \pm 0.479
Item14	0.851 \pm 0.358	0.987 \pm 0.113
MI-consistent attitudes	2.297 \pm 1.190	2.744 \pm 1.145
Item4	0.311 \pm 0.466	0.487 \pm 0.503
Item8	0.541 \pm 0.502	0.718 \pm 0.453
Item10	0.622 \pm 0.488	0.680 \pm 0.470
Item13	0.811 \pm 0.394	0.859 \pm 0.350
Checklist	2.892 \pm 1.740	4.078 \pm 1.204
Correct	2.892 \pm 1.740	4.078 \pm 1.204
Item16	0.757 \pm 0.432	0.910 \pm 0.288
Item18	0.743 \pm 0.440	0.872 \pm 0.336
Item21	0.270 \pm 0.447	0.667 \pm 0.474
Item24	0.514 \pm 0.503	0.821 \pm 0.386
Item29	0.581 \pm 0.497	0.808 \pm 0.397
Incorrect	2.135 \pm 2.383	1.577 \pm 2.452
Item15	0.230 \pm 0.424	0.154 \pm 0.363
Item17	0.581 \pm 0.497	0.256 \pm 0.439
Item19	0.284 \pm 0.454	0.128 \pm 0.336
Item20	0.095 \pm 0.295	0.064 \pm 0.247
Item22	0.176 \pm 0.383	0.103 \pm 0.305
Item23	0.054 \pm 0.228	0.077 \pm 0.268
Item25	0.054 \pm 0.228	0.051 \pm 0.222
Item26	0.189 \pm 0.394	0.128 \pm 0.336
Item27	0.203 \pm 0.405	0.256 \pm 0.439
Item28	0.176 \pm 0.383	0.141 \pm 0.350
Total MIKAT Score (#/19)	11.027 \pm 4.210	15.256 \pm 3.064

change) achieved the equal highest means at post-training (.923), thus suggesting the participants found these items in particular less challenging.

Internal consistency. The evaluation of internal consistency was computed using Cronbach's α coefficients along with component and subcomponent correlations. The coefficients for the total MIKAT, the components, and subcomponents are displayed in Table

3. The total score MIKAT coefficient at pre-training was adequate ($\alpha = .765$). Although relatively close, the post-training coefficient failed to reach the minimum acceptable value for newer psychological measures (.70; Lance et al., 2006). Both the total coefficients were lower than the previously reported internal consistency for the MIKAT ($\alpha = .84$; Doran et al., 2011a).

All component coefficients, with the exception of post-training true-false were adequate ($>.70$), however the subcomponents were variable. The items for MI-consistent attitudes were relatively poor across time points ($\alpha = .50 - .52$) whereas myth and correct options on the checklist decreased from pre- to post-training (pre- $>.70$; post- $<.70$). Additionally, incorrect responses on the checklist increased across time points ($\alpha = .77 - .80$).

Table 3. Internal Consistency of the Total, Component, and Subcomponent MIKAT Scores for Pre- and Post-Training

	Internal Consistency	
	Pre-	Post-
True – False	.738	.609
Myth	.832	.690
MI Consistent	.502	.518
Checklist	.788	.751
Correct	.799	.611
Incorrect	.777	.800
Total Score	.765	.665

Component and subcomponent scale correlations for the MIKAT are displayed in Table 4. This table shows that the true-false and checklist components were adequately correlated, however correlations among the subcomponents were relatively low and all decreased from pre- to post-training. Myth and correct checklist choices exhibited the largest subcomponent correlations across time points. MI consistent principles exhibited some of the lowest correlations along with the most variation across time points. The extent of the variation indicated that the items in this subcomponent were not adequately robust. The low

relationships among the remainder of the subcomponents implied that they were measuring different constructs; this paired with the adequate Cronbach's α coefficients across these subcomponents indicated that analysis of the total MIKAT scores would not be sufficient.

Table 4. Scale correlations of the MIKAT Components and Subcomponents for Pre- and Post-Training

	Pre- (r_c)	Post- (r_c)
True-False		
Checklist	.519	.434
Myth		
MI Consistent	-.199	-.007
Correct Choice	.433	.393
Incorrect Choice	-.138	-.105
MI consistent		
Correct Choice	.254	.184
Incorrect Choice	.198	-.072
Correct Choice		
Incorrect Choice	.258	.165

Factor Analysis. To further investigate the appropriate form of analysis for the MIKAT, two PCAs (based on the subscales) were conducted, one for pre-training and one for post-training. Due to the unacceptable Cronbach's α coefficients, correlations, and amount of variation of the MI-consistent attitudes subcomponent, these items were removed prior to the PCA and thus excluded from further analysis. It must be acknowledged that with the exclusion of these items MI attitudes was no longer a component of the measure, therefore the MIKAT was solely concerned with the knowledge of MI.

Table 5 displays component loadings generated by the PCA. It is clear by the size of the myth and correct choice loadings onto one component and that of the incorrect choice onto another component that the MIKAT loaded onto two components across pre- and post-training; myth and correct checklist choices loaded onto the first and incorrect choices, the second. There was some cross loading by correct checklist choices, however in comparison to the loadings of the other subcomponents, it was relatively minimal.

Table 5. Component Loadings of the MIKAT

		Component	
		1	2
Pre-	Myth	.886	-.288
	Correct Choice	.800	.379
	Incorrect Choice	-.023	.958
Post-	Myth	.852	.240
	Correct Choice	.825	.313
	Incorrect Choice	.024	.962

The amount of explained variance reinforced the two-component model of the MIKAT; 85.70% of the variance was explained by the two components and at post-training 82.38% was explained. Therefore two components that account for the three remaining subcomponents explained over 80% of the variance of MI knowledge at both pre- and post-training.

The relatively adequate Cronbach's α coefficients across the subcomponents (with the exclusion of MI-consistent attitudes) along with low scale correlations indicated the analysis of the MIKAT as a single measure was not appropriate. This was supported by the presence of a two-component model identified by a PCA to explain over 80% of the variance in MI knowledge. Therefore it was decided to analyse the MIKAT as having two separate components; the myth scores and checklist choice scores comprised one component and the incorrect checklist choice scores comprised the second.

Sensitivity to training effects. With the exclusion of the MI-consistent attitudes subcomponent, the scores of myth and correct checklist choice composed an updated total MIKAT score, this time out of 15. The updated means and standard deviations for this total score increased from pre-training (8.73 ± 4.01) to post-training (11.56 ± 2.71). Additionally, the incorrect choice scores decreased from pre- to post-training ($2.12 \pm 2.38 - 1.58 \pm 2.45$).

A series of Mann-Whitney U tests were run to determine whether there were significant differences in the total and incorrect checklist choice scores of the MIKAT from pre- to post-training. Eighteen participants had incomplete scores on the MIKAT, eleven did not complete a pre-test and seven did not complete a post-test. According to Little's MCAR test (Little, 1988) the missing data were not missing completely at random. The analyses were run once with the incomplete cases excluded and once using multiple imputations (see method section for description).

With a sample size of $N = 74$ for the pre-training and $N = 78$ for post-training the analyses had 85% power assuming a medium effect size ($d = .50$). The total MIKAT mean ranks significantly increased from pre- (56.66) to post-training (93.43), $U = 1,565.50$, $z = -2.30$, $p < .001$, $d = .83$. With regards to the incorrect checklist choice scores, a significant decrease was observed in the mean ranks from pre- (84.60) to post-training (68.81), $U = 2,286.5$, $z = -2.30$, $p = .02$, $d = .22$. Collectively, these results indicate that from pre- to post-training, performance on the MIKAT significantly improved through the increase in myth and correct choice scores and the decrease in incorrect choice scores.

The analysis, run again using multiple imputations, revealed results comparable to the original data. A power analysis was run for the complete sample size ($N = 85$) and revealed 88% power. The total score mean ranks again significantly increased (85.06–105.94), $U = 1,875.5$, $z = -5.431$, $p < .001$. There was also a significant decrease in the mean ranks for incorrect checklist scores (95.90–75.10), $U = 2,728.5$, $z = -2.831$, $p = .004$.

Pearson correlation was then used to determine the extent of the relationship between the total MIKAT pre- and post-training scores. There was no relationship between the scores ($r = .02$, $p = .867$). The Gower coefficient of agreement (see method section for details) was then used to assess the agreement between the training scores. This analysis measured the

agreement (or similarity) in change of score from pre- to post-training regardless of location. There was a significantly high coefficient of agreement (.76, $p < .001$) indicating that, irrespective of location, the changes in scores from pre- to post-training were 76% similar despite the absence of monotonicity exhibited using Pearson correlation. The prior Mann-Whitney U tests determined that this was a positive change, i.e., scores increased from pre- to post-training.

The VASE-R

Inter-rater reliability. All measure scoring was completed by the single rater, however, a subset of 19 (20%) tests were also scored by the member of MINT to ensure an adequate level of inter-rater reliability was achieved. The intra-class correlations (ICCs) were interpreted with the use of published guidelines (Cicchetti, 1994). The result for the ratings of the total VASE-R scores was in the excellent range ($ICC = .935$), indicating a high level of agreement; the VASE-R was rated similarly across both authors.

Further ICCs were computed for each of the five VASE-R subscales: reflective listening (RL), responding to resistance (RR), summarizing (S), eliciting change talk (ECT), and developing discrepancy (DD). There was some variation amongst the subscales, with the strongest reliability being for the DD scale, having perfect agreement ($ICC = 1.000$). This was closely followed by RL ($ICC = .952$), RR ($ICC = .868$), ECT ($ICC = .820$), and S ($ICC = .762$). Despite the variation, all remained in the excellent range (.75 – 1.00, Cicchetti, 1994). These results demonstrated that although the high level of agreement reflected in the total score coefficient was not consistent across all five subscales variation remained in the excellent range.

Rosengren et al. (2005) recommended 90% agreement across raters however, he did not specify whether this recommendation regarded the total score or was across the individual

subscales. It was believed that the guidelines concerned the agreeability of the total VASE-R score as it was likely recommendations across subscales may have varied. Therefore, the inter-rater reliabilities were deemed suitable for use in the present analyses.

Descriptive statistics. Table 6 displays the means and standard deviations for the individual items, five subscales, and total VASE-R scores for pre- and post-training. Results reflect variation in difficulty across the items and subscales. For example, it is clear that the participants found the items in the S and ECT subscales more challenging, with consistently lower scores (1.13–1.64, 1.80–2.64, respectively) compared with RL, RR, and DD (4.26–4.94, 4.44–6.03, 3.33–3.80, respectively). All items and subscales increased from pre- to post-training with the RR and ECT scales exhibiting the largest increases.

Table 6. Means and Standard Deviations for VASE-R Items, Subscales, and Total Scores for Pre- and Post-Training

	Pre- (N=45)	Post- (N=39)
	M ± SD	M ± SD
Reflective Listening (RL)	4.267 ± 2.349	4.949 ± 2.470
Item 1	1.222 ± 0.850	1.436 ± 0.821
Item 2	1.089 ± 0.821	1.256 ± 0.785
Item 7	0.956 ± 0.825	1.128 ± 0.833
Item 9	1.067 ± 0.837	1.154 ± 0.814
Responding to Resistance (RR)	4.444 ± 2.546	6.026 ± 2.611
Item 3	1.067 ± 0.751	1.487 ± 0.756
Item 8	0.822 ± 0.806	0.974 ± 0.811
Item 13	0.889 ± 0.745	1.256 ± 0.715
Item 14	0.867 ± 0.757	1.128 ± 0.833
Item 15	0.756 ± 0.773	1.205 ± 0.833
Summarizing (S)	1.133 ± 1.160	1.641 ± 1.203
Item 4	0.578 ± 0.657	0.667 ± 0.663
Item 10	0.267 ± 0.495	0.564 ± 0.552
Item 16	0.267 ± 0.447	0.410 ± 0.549
Eliciting Change Talk (ECT)	1.800 ± 1.358	2.641 ± 1.630
Item 5	0.578 ± 0.583	0.846 ± 0.586
Item 11	0.667 ± 0.707	1.077 ± 0.774
Item 17	0.533 ± 0.588	0.667 ± 0.701
Developing Discrepancy (DD)	3.333 ± 2.205	3.795 ± 1.780
Item 6	1.089 ± 0.996	1.128 ± 1.004
Item 12	1.044 ± 0.999	1.385 ± 0.907
Item 18	1.200 ± 0.991	1.333 ± 0.955
Total VASE-R score	14.933 ± 6.820	19.051 ± 7.189

Internal consistency. The evaluation of internal consistency was computed using Cronbach's α coefficients and the correlations between each of the five scales. The coefficients for the total VASE-R along with the subscales for pre- and post-training are displayed in Table 7. Internal consistency for the total scores at both pre- and post-training were adequate ($\alpha=.816$, $\alpha=.829$, respectively) and comparable to that of Rosengren et al. (2008, $\alpha=.85$). With regards to the subscales there was variation and only the coefficients of the RL and RR subscales neared the minimum acceptable value for newer psychological measures ($\alpha=.70$; Lance et al., 2006) for both pre- and post-training. However, all subscales, with the exception of S, had at least one coefficient (pre- or post-) which approximated that of Rosengren et al. (2008).

Table 7. Internal Consistency of the Total and Subscale VASE-R Scores for Pre- and Post-Training

	Internal Consistency	
	Pre-	Post-
RL	.666	.753
RR	.678	.660
S	.510	.417
ECT	.439	.606
DD	.583	.250
Total Score	.816	.829

The Cronbach's α coefficients were then used to disattenuate the scale correlations, displayed in Table 8. There was again much variation across the subscales with several correlations being unacceptably low. Those that exceeded 1.0 suggest the scores were not normally distributed, which would create problems for further analyses.

With the information from the evaluation of internal consistency, preliminary assumptions were made about the form in which to analyse the VASE-R. No coefficients achieved the minimum acceptable value and those of the S, ECT, and DD subscales were

particularly low. Combined with the relatively variable scale correlations, this indicated that it would not be appropriate to divide the VASE-R into subscales for analysis.

Table 8. Disattenuated scale correlations of the VASE-R Subscales for Pre- and Post-Training

		Pre- (r_c)	Post- (r_c)
RL	RR	.496	.562
	S	1.036	1.110
	ECT	.453	.816
	DD	.303	.519
RR	S	.738	.564
	ECT	.988	.718
	DD	.639	1.494
S	ECT	.921	1.015
	DD	.814	.424
ECT	DD	.585	1.028

Factor analysis. As a further investigation into the appropriate form in which to analyse the VASE-R, a parallel analysis for both pre- and post-training was conducted. Table 9 lists the five actual eigenvalues drawn from the present sample along with the mean eigenvalues and those at the 95th percentile of random data (to be retained for the PCA, an actual eigenvalue must be higher than the 95th percentile, see method section for further explanation). An examination of Table 9 indicates that only the first eigenvalue for both pre- and post-training was greater than the 95th percentile value of the random data, thus suggesting a single component at each time point is to be retained for analysis.

A PCA was then conducted using the one component recommended by the parallel analysis to determine the amount of variance explained by this model. The PCA determined that at pre-training, 51.8% and at post-training 53.93% of the variance was explained by one

component. Therefore this one component that accounts for the five subscales explained over 50% of the variance of MI skill both pre- and post-training.

Table 9. Eigenvalues Generated from the Parallel Analysis: the Actual Eigenvalues Correspond to the VASE-R whereas the Mean and 95th Percentile were Generated at Random.

	Actual Eigenvalue	Mean Eigenvalue	95 th Percentile Eigenvalue
Pre-training	2.589	1.438	1.658
	0.900	1.172	1.312
	0.734	0.981	1.084
	0.471	0.797	0.918
	0.305	0.610	0.749
Post-training	2.696	1.478	1.710
	1.120	1.187	1.329
	0.462	0.978	1.092
	0.375	0.784	0.911
	0.347	0.572	0.726

Having adequate Cronbach's α coefficients for the total scores and low, variable coefficients across the subscales primarily indicated that analysis of the VASE-R as a single measure would be more appropriate. This was supported by the single significant component identified by the parallel analysis and the adequate amount of variance explained by this component, as determined by the PCA. Therefore it was decided to analyse the VASE-R using only the total scores.

Sensitivity to training effects. A paired samples t-test was used to determine whether there was a significant difference between pre- and post-training scores on the VASE-R. There was a significant amount of missing data among the sample: eight participants had incomplete scores; seven did not complete a post-test and one did not complete a pre-test. The data were not missing at random according to Little's MCAR test (Little, 1988). As with the previous MIKAT data the analysis was run once with the incomplete cases excluded and once using multiple imputations.

With a sample size of $N = 38$ the analyses had 85% power to detect a medium effect size ($d = .50$). When investigating the total VASE-R scores, a preliminary assumption check indicated that data were normally distributed as assessed by a Shapiro Wilk test ($p=.46$) and that no outliers were detected. Participants improved their scores from pre- to post training (14.93 ± 6.82 – 19.05 ± 7.19), a statistically significant increase of 4.03 (95% CI), $t(37)=6.409$, $p<.001$, $d = 0.59$.

Because the VASE-R scores were paired it was possible to create a discrepancy histogram that displayed the change scores for each participant (Figure 1). This figure provides a visualisation of how many participants' scores changed by certain numbers of points.

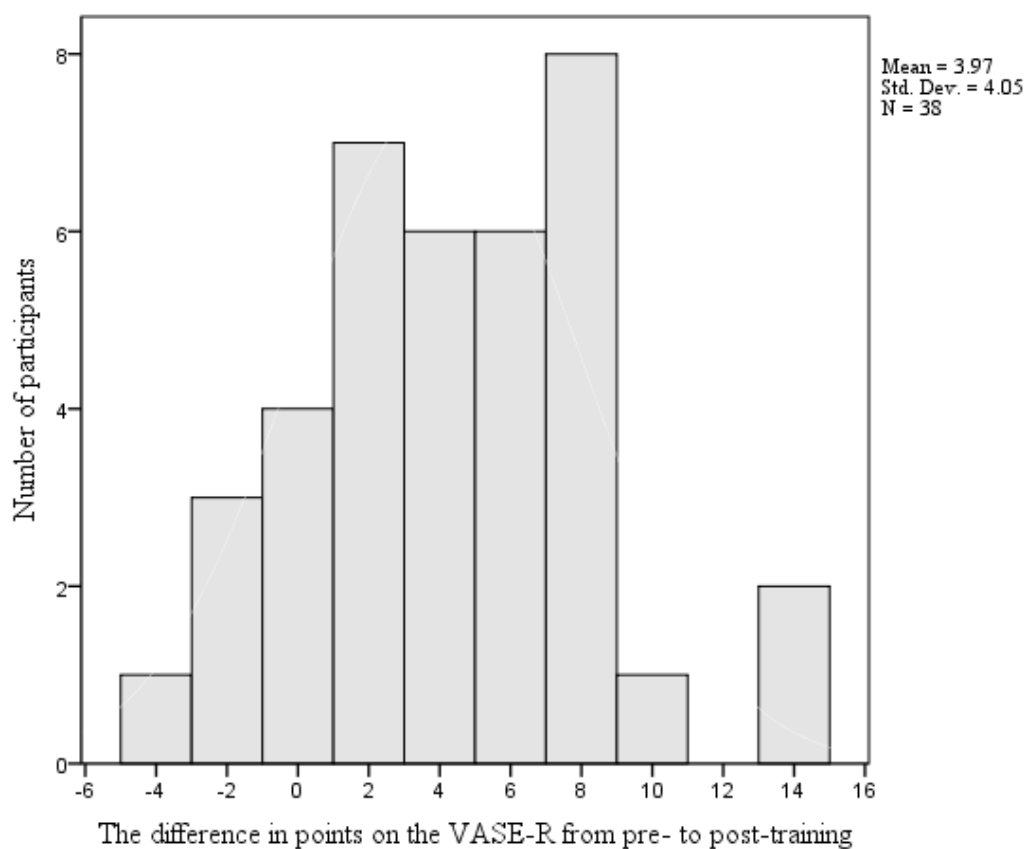


Figure 1. Discrepancy histogram displaying the similarities in VASE-R point changes from pre- to post-training. VASE-R = Video Assessment of Simulated Encounters–Revised.

The analysis was run again using multiple imputations and revealed little difference to the original data. A power analysis was run for the complete sample size ($N = 46$) and revealed 91% power. Again participants' scores improved from pre- to post-training (14.791 ± 6.828 – 18.757 ± 7.180), a statistically significant difference of 3.96 (95% CI), $t(45) = 6.509$, $p < .001$, $d = 0.57$.

Finally, Pearson correlation was used to determine the extent of the relationship between the total VASE-R pre- and post-training scores. There was a high correlation between the sets of scores ($r = .83$). The correlation was then disattenuated for measurement error, increasing the value to demonstrate perfect linearity ($r_c = 1.009$). The Gower coefficient of agreement was then used to assess the agreement between training scores. Again, this analysis measured the agreement (or similarity) in change from pre- to post-training across the participants, regardless of location. There was a significantly high coefficient of agreement (.82, $p < .001$) indicating that, irrespective of location, the changes in scores from pre- to post- training were 82% similar. The prior t-tests indicated that these changes were in a positive direction, scores increased from pre- to post-training. Referring back to Figure 1, the visualisation of participant change scores also allowed for further insight into the agreement across changes. Inspection of this figure revealed that, although four participant's scores (10.52%) decreased from pre- to post-training and three more (7.90%) exhibited an increase of >10 points, the remainder of the participant's (81.57%) score changes were <10 points.

Discussion

Because MI continues to be utilised as a method to assist in behaviour change, the importance of effective training in this method is imperative. Therefore, the need for reliable measures that assess skill attainment in MI remains important. There are a handful of MI measures assessing skill and adherence (Leffingwell, 2006; Miller et al., 2008; Moyers et al.,

2005; Rosengren et al., 2008); however the most reliable are also the most labour intensive.

The goal of the present research was to investigate whether the MIKAT and the VASE-R were reliable and valid measures of MI knowledge and skill. Data was used from a Child Youth and Family services (CYFs) MI training programme in order to evaluate the measures.

The results from the present research indicated that although the MIKAT and VASE-R effectively demonstrated sensitivity to training effects, it appeared both measures require refinement and even restructuring to increase their validity and make them more consistent with the most recent MI developments.

Summary of Findings

The present research provides further insight into the psychometric properties of the MIKAT and the VASE-R. Several of the analyses were replications of those conducted in previous investigations and, although differing slightly, were relatively consistent and supported these earlier findings. Specifically, the pre-training internal consistency for the total MIKAT score was adequate and comparable, albeit slightly lower, to the only other internal consistency reported for the measure (Doran et al. 2011b). With regards to the VASE-R, the internal consistency for all subscales, with the exception of the summary (S) subscale, approximated that of Rosengren et al. (2008) at pre-training, post-training or both. Additionally, inter-rater reliability of the VASE-R was comparable with Rosengren et al. with the scores in the present research being marginally higher.

The MIKAT and the VASE-R also demonstrated sensitivity to effects of training, another outcome that was consistent with previous research (MIKAT, Doran et al., 2011b; Hohman et al., 2009; Leffingwell, 2006; VASE-R, Bell & Cole, 2008, Doran et al. 2011a; Rosengren et al., 2008; see the Introduction for details of these studies). Additionally, increases across the scores were relatively consistent for both measures: seventy-six per cent

of participants that completed the MIKAT had similar increases in their scores whereas 82% of VASE-R participants' scores increased similarly (<10 points).

Specifically, in the present study the MIKAT scores significantly increased from pre-training (8.73) to post-training (11.56). Although both Leffingwell (2006) and Doran et al. (2011b) had results consistent with the present study, only Leffingwell's sample was of a comparable size ($N=71$). Despite having a slightly smaller sample size than the present research ($N=74$) it is unlikely Leffingwell experienced problems with reduced power because paired samples t-tests were used. The use of t-tests indicated that the pairing of participant scores was possible, unlike in the present study. The required sample size for 95% power in a paired samples t-test is $N=54$, therefore Leffingwell had an adequate sample. Doran et al. also used a paired samples t-test, however had a much greater sample size ($N=1,522$) therefore it is unlikely power was a problem in that study.

As with the present research, Leffingwell (2006) and Doran et al. (2011b) reported significant increases in MIKAT performances (both $p<.001$), however neither reported the amount of points their samples increased by. Also, Leffingwell went into more detail reporting both the increases in the MI-consistent attitudes and the correct choices made on the checklist separately along with the decreases in the selection of addiction myths and incorrect choices ; all specific increases and decreases were significant ($p<.001$).

To assess the degree of change in MI knowledge shown by the present sample with previous studies, effect sizes were compared. However, this comparison was difficult due to each study interpreting the MIKAT scores differently. Additionally, pre- and post-training scores could not be paired in the present study which would have reduced the effect sizes. Leffingwell (2006), the developer of the MIKAT, evaluated the measure at each subcomponent (outlined earlier), therefore generating four effect sizes. Very large effects

were reported by Leffingwell ranging from $d = 1.07$ for the decrease in incorrect choice on the checklist to $d = 1.95$ for the correct identification of the addiction myths as false. Doran et al. (2011b) simply analysed the total score of the MIKAT, also reporting a large effect size ($d = .96$). In the present study the MI-consistent attitudes subcomponent had been removed (explained further in subsequent sections) and the MIKAT was analysed as two components, thus two effect sizes were computed. A large effect was identified for the myth and correct choice scores ($d = .83$), however, a small effect was reported for the decrease in incorrect choice items ($d = .22$). Although the effect size from the present study was marginally lower than other research, it can still be noted that with regards to the total MIKAT scores the research so far has reported large effects. However, with regards to the decrease in incorrect choice scores, the effect calculated in the present study was much lower than the only other report on the subscale (Leffingwell, 2006). The inability to pair pre- and post-training scores must again be acknowledged as it would have contributed to the smaller effect sizes of the present study.

With regards to the VASE-R, scores significantly increased pre-training (14.93) to post-training (19.05) in the present study. Previous research utilising the VASE-R also saw significant VASE-R increases with their participants (Bell & Cole, 2008; Doran et al., 2011a; Rosengren et al., 2008), however only one of these studies used a paired samples t-test, like in the present study (Bell & Cole, 2008). Doran et al. (2011a) used repeated measures analysis of variance (ANOVA) measuring at two time points (pre- and post-training) whereas Rosengren et al. (2008) used a repeated measures multivariate analysis of variance (MANOVA) measuring at pre- and post-training, along with a three-month follow-up. Sample sizes for these studies were also larger than that of the present research ($N = 38$). Bell and Cole (2008) had a sample size of $N = 53$, one less than what is required for 95% power in

a paired samples t-test and both Doran et al. and Rosengren et al. had much larger samples of $N=222$ and $N=144$ respectively.

Again, all the previous studies reported significant increases in total VASE-R scores (all $p<.001$). The amount of points participants increased by differed slightly across the studies. Because only the first vignette of the VASE-R was used in Bell and Cole's (2008) research, the increase from 7.04 at pre-training to 9.47 and post-training was not comparable to the other research. The participants in the remaining two studies (Doran et al., 2011a; Rosengren et al., 2008) performed better than the present study at both pre- and post-training and the increases were slightly larger in these previous studies. Doran et al.'s participants increased from 17.70 points at pre-training to 22.00 points at post-training. Rosengren et al.'s participants performed better still, they started with 18.21 and increased to 24.13 after training. Additionally, Rosengren et al. reported a three-month follow-up which demonstrated a small decline to 23.08. However, this decrease was not significant and participant scores remained higher than in the other studies.

Effect sizes were also compared across the studies to assess the degree of change in MI skill. The effect sizes reported by the previous research were consistently larger than those from the present study. Specifically, Bell and Cole (2008) reported a very large effect ($d=1.54$) for their total VASE-R score; although only the first vignette was utilised in this research, the effect can be viewed as an indication of how the participants would have performed if administered the entire measure. Doran et al. (2011a) and Rosengren et al. (2008) also reported large effect sizes with their participants ($\eta^2=.53$ and $.50$ respectively). These effects were significantly larger than the medium effect identified in the present study ($d=.59$).

Additional analyses such as parallel analysis and principal components analysis were performed in the present study. These analyses had not yet been implemented with the MIKAT and VASE-R and identified the most appropriate form in which to use the measures. In doing so, several shortcomings of each measure were also uncovered. The subcomponents of the MIKAT were revealed to measure different aspects of MI knowledge. Therefore, the total score of the measure was deemed unusable for analysis; the subcomponents needed to be measured separately.

During this investigation, issues arose with the items measuring MI-consistent attitudes. It appeared that the items comprising this subcomponent were either too broad or they were providing an inaccurate measure. It was also likely that there were too few items (four) to reliably measure the principles they were intended to measure. This necessitated the removal of these particular items which then narrowed the focus of the MIKAT to just MI knowledge. Moreover, the exclusion of the items also rendered the measure incomplete because the MI-consistent attitudes were the items to be marked *true* in the true–false component and without them, only the addiction myths (false statements) remained.

The removal of the MI-consistent attitudes subcomponent also had implications for the interpretation of training outcomes. Without this subcomponent the MIKAT was not an accurate measure of MI attitudes, therefore the results from the present research reflect MI knowledge and the ability to correctly identify addiction myths.

With regards to the VASE-R there was much overlap in MI skill across the subcomponents, thus indicating that the most appropriate way in which to analyse the measure was through the total score. This outcome determined that the VASE-R can only provide a single overall measure of skill attainment, therefore limiting the interpretation of

outcomes. It is expected, given the development of the subscales, that the VASE-R was intended to provide more specific outcomes.

Additional to the analyses, it was noted that some of the content of the MIKAT could be outdated. The measure has not been subjected to any revisions since its development in 2006 and the checklist component of the measure consisted of the core MI principles that, since the recent developments made in MI (Miller & Rollnick, 2012), no longer exist. These principles were not omitted from MI however; the method underwent rigorous revision resulting in the principles being incorporated into the structure and therefore are not acknowledged separately. Although the items included in the checklist remain consistent with MI, they are no longer principles.

Finally, it also appears that the content of the MIKAT may be too specific to the field of substance abuse and therefore could limit its use with other populations. CYFs' training participants provided feedback on the MIKAT and reported difficulties with particular myth items that were not completely relevant to them. Although some of the vocabulary in the measure had been altered in an attempt to generalize the statements (see Appendix E for the MIKAT modified for CYFs' participants), several items such as "Substance abusers are generally incapable of making sound decisions in their current state of addiction", and "Addicts and alcoholics are not capable of exerting control over their substance use behaviour" remained specific to the field of substance abuse and were not considered applicable by all participants.

Recommendations

In light of the findings from the present study, it is recommended that the MIKAT and the VASE-R undergo revision and possible restructure. Specific recommendations are made for each measure and are as follows.

The MIKAT. As stated previously, the subcomponents of the MIKAT measure different aspects of MI knowledge and cannot be analysed together as a single measure. The MI-consistent attitudes subcomponent also had serious problems with validity which affected the measure as a whole and resulted in its removal.

The simplest alteration that could possibly improve the validity of the MIKAT would be to include more MI-consistent attitude items, thus making the subscale larger. The lack of items in this subscale may have contributed to the weak validity. The addition of further attitude statements, along with the revision of the existing items, may help to strengthen this subcomponent. The myth items could assist in the construction of new attitude statements as several of the myths can be viewed as being contradictions of MI views. For example, a statement in accordance with MI such as “The use of labelling (for example: using the term alcoholic to describe a person with alcohol problems) has been known to increase resistance in clients and is best avoided” could be derived from the myth “Clients must accept their problem (for example: I am an alcoholic/addict) before they can get help.”

Another possibility could be to take a different approach to measuring MI attitudes by removing the true–false component completely. It is unclear what format would be the best replacement for the component; the most desirable format would be one that is equally as straightforward so the measure does not become overly time consuming. A possibility could be adopting a format similar to the checklist measuring MI knowledge. However, rather than selecting correct options from a list, the statements could be arranged in pairs and the responder selects the statement that is in accordance with MI beliefs. An example of this format is displayed in Figure 2.

MOTIVATIONAL INTERVIEWING QUIZ

The following pairs of statements contain one option that is factually true and one that is false or consistent with ("true") or inconsistent with ("false") a motivational interviewing approach. Indicate your response by checking the necessary box as true.

1. External pressure and consequences is the only way to make clients change. ☐
2. Demonstrating empathy towards a client along with forming a collaborative relationship can reduce client resistance and increase the possibility of change. ☐

Figure 2. An example of the recommended paired-items format to replace the true–false component.

This format may not require the two statements to be in direct contradiction. Pairing statements from the original MIKAT could develop new items as shown in Figure 3.

However, the fact that the items are not strictly paired (i.e., are not reflecting two sides of the same issue) could lead to complications with the format. Therefore, a comparative validity study is recommended to evaluate the approach if it is considered.

Despite the fact that with this format the MIKAT would essentially consist of two checklists, they would still exhibit subtle style differences along with differences in content. That is, the first would measure MI-consistent attitudes and the second, MI knowledge. This format has several benefits including having little effect on the amount of time the MIKAT would take to complete. Additionally, having to select the correct answer in the paired-items component means that only MI-consistent attitudes are measured whereas the true–false component measures both MI-consistent attitudes and addiction myths. Therefore, due to the specificity of measurement, stronger validity may be observed with this format.

With regards to the checklist, the items in this component require updating. Because the existing items are still relevant to MI it is recommended that they remain within the measure but that the four core processes (outlined in the Introduction) be added to keep the

component consistent with the recent revisions of MI. The four core processes of MI are what comprise the structure of the method and consist of engaging, focusing, evoking, and planning (refer to the structure of Motivational Interviewing section of the Introduction for further details). Statements with the processes in mind could be added to the former MI principles to provide an adequately updated checklist. The following are examples of potential statements that could be included: “Engage in a collaborative relationship with the client.”; “Focus the therapy on a possible area of change that is important to the client.”; “Evoke change talk.”; “Assist in the formation of a change plan (if the client desires).”

MOTIVATIONAL INTERVIEWING QUIZ

The following pairs of statements contain one option that is factually true and one that is false or consistent with (“true”) or inconsistent with (“false”) a motivational interviewing approach. Indicate your response by checking the necessary box as true.

1. Counsellors should emphasize personal choice over clients' behaviours, including substance abuse.. ☐
2. Addicts and alcoholics are not capable of exerting control over their substance use behaviour. ☐

Figure 3. A further example of the recommended paired-items format, this time using items from the original MIKAT.

Along with the addition of the processes to the checklist, the statement introducing the component must be altered due to the fact that there are no longer distinct principles of MI. By simply changing the wording from “Which of the following are principles of a Motivational Interviewing approach?” to “Which of the following are *consistent with* a Motivational Interviewing approach?” the statement then includes the processes whilst keeping the former principles as MI-consistent statements.

As a final recommendation, the MIKAT needs to consist of items that are applicable to practitioners working in any field that has incorporated MI including those who work with

youth. In the current MIKAT the MI-consistent attitudes are already very general as they are consistent with a view rather than having a focus on a particular client group. However, myth items need to either be broadened to apply to all fields or several versions of the MIKAT must be developed with the more specific items altered to best reflect a situation with a particular type of client. The latter option appears more feasible due to the diversity of the main fields that currently utilise MI: substance abuse, criminal justice, health systems, and youth (which could fall under any of the other fields). Developing comparable items specific to each field may be less labour-intensive than composing a single item that applies to all fields. However, not all items would need to be altered. For example, “Therapists’ expectancies for their client’s abilities to change have no effect upon whether change occurs” is more general, with a focus on the practitioner. Therefore, this could remain consistent across several versions of the MIKAT. As with the development of new MI-consistent items, original myths can assist in the development of new ones, along with keeping a consistent context across items. For example, “Substance users need to ‘hit rock bottom’ before they can change” can be altered to suit other fields. For adult offenders, the statement could be “Offenders need to be punished before they will change.” A potential youth offender item could be “Youth offenders need to be frightened with the prospect of incarceration to motivate them to change.” An item regarding individuals with health related issues could be “Patients with life threatening illnesses need to be frightened by the risks before they will be motivated enough to make lifestyle changes.”

To conclude, the MIKAT was identified to have several limitations as a result of the present evaluation (outlined in previous sections). These shortcomings were traced to the validity issues of the MI-consistent attitudes component and it is therefore recommended that more items be added to the measure as a way of improving the true–false component, along with updating the checklist, before submitting the measure to further psychometric

investigation. This would determine whether or not the validity of the measure has improved or whether a replacement of the true–false component is required.

The VASE-R. The major shortcoming of the VASE-R was that because only the total score can be used to analyse the measure, results will lack specificity: it is expected that the VASE-R should provide more informative measures of MI skill in terms of particular strengths and weaknesses. One possibility that could improve the internal consistency of the measure and therefore clearly divide the items into valid subscales could be to alter the questions in the answer booklet so that they are more specific. This would ensure that the responder has a good idea of what is being asked of them.

In the current VASE-R answer booklet there are five prompts (one representing each of the subscales) that are asked at least once within each vignette, inviting a response (see Appendix C for the prompts used in the VASE-R). However, with the exception of the reflective listening (RL) items, they do not convey exactly what is required from the responder; it is the responder’s job to identify what MI skill is being called for. Although this expectation is consistent with the goal of the VASE-R—to create a situation approximating a real client experience—it may be too difficult for more novice trainees. It is therefore recommended that the prompts inviting responses for the responding to resistance (RR), summary (S), eliciting change talk (ECT), and developing discrepancy (DD) subscales are modified to more specifically prompt the responder’s answers. As previously mentioned, the statements for the RL subscale appear suitable (“Write a response that indicated you are listening.”); however, the other prompts are too vague. The following are possible alternatives for the other four subscales: for RR, “How might you respond to the resistance exhibited by (client)?”; for ECT, “Write what you would say to (client) that might encourage the evocation of change talk (e.g. express concerns and/or recognition of problems, intention and/or optimism about prospective changes).”; for S, “Write a summary that you might say to

(client) that touches on the things that you think are most important (note that a summary requires two ideas and includes change talk and ambivalence).”; and for DD, “There are many different directions to explore with (client), select the statement you think might help (client) identify a discrepancy in their current situation and may increase their motivation for change; indicate why you chose that one in the space below.” Such alterations would be subtle and would not change the structure of the prompts, however, they would make it clear which particular MI skill was being evoked.

It is acknowledged that the suggested prompts remove authenticity from the measure because the responder has more guidance. This differs from VASE-R’s overall goal of simulating everyday encounters with clients. However, it is expected that more novice trainees in MI would need this specificity in order to be able to demonstrate their acquired MI skills. However, this may not be the case for more experienced trainees (those in the feedback and coaching phase of training) or practitioners who have a greater understanding of what is being asked of them. Therefore, two VASE-R measures comprising different prompts could be created; one that includes specific prompts such as those offered above and which is more suitable for novice MI trainees and an alternative that retains the vague, more challenging prompts and is suited to more experienced trainees or practitioners.

Internal consistency of the original VASE-R has not been assessed with a population of trainees with more MI experience. If validity is improved with this trainee population and a factor analysis identifies the subscales, then the original VASE-R could be retained for use with trainees with a higher level of competency. A second VASE-R measure could then be developed for novice trainees who require more guidance.

If making the prompts of the VASE-R more specific does not improve validity at the subscale level, modification of the subscales could be another possibility. As with the

MIKAT, the VASE-R includes some of the former MI principles that existed prior to the revision of the method (i.e., responding to resistance and developing discrepancy). Modifying the subscales to be more consistent with the current structure of MI would mean replacing some of the outdated subscales or combining them to make new subscales. The revised subscales could follow the structure of MI: engaging, focusing, evoking, and planning. For example, some reflective listening, responding to resistance, and summary items could be combined to form an engaging subscale. The same types of items would be relevant in the focusing subscale; however the prompts would need to be made contextually diverse so the responder is aware of what process they are expected to be in. Eliciting change talk items could comprise the evoking subscale, using the more specified prompt to ensure the responder is aware that a response that would increase change talk is required. Although not all clients need the planning process of MI, it is expected that practice in this area may be beneficial; it is therefore recommended that items regarding planning are added to the VASE-R. Modifying the subscales of the VASE-R would ultimately lead to a revision of the measure in its entirety, updating the content to be more consistent with the present MI structure. It may be found that combining items that were previously in separate subscales, along with the modification and addition of other items, that the VASE-R may have the potential to give more specific insight into trainee progress in MI skill attainment.

To conclude, the present research identified that only the total score of the VASE-R accurately measures MI skill attainment. The subscales do not separately measure different skills therefore, in order to achieve more specific outcomes from the test, modifications were recommended. Although MI has undergone recent revision, the content of the original VASE-R remains valid; the techniques that require demonstration are still major components of MI. It was for that reason the first recommendation was suggested to be the alteration of the prompts to be more specific. This task would be less taxing than modifying the entire

measure in order to create new subscales. However, given the recent revisions of MI, a full modification of the measure might be the most appropriate course of action. Additionally, the prompts could be made more specific in the current version of the VASE-R which would result in two separate measures of MI skill: the original VASE-R with more specific prompts which would be suitable for newer trainees; and a completely modified test that is consistent with the current structure of MI and requires the acknowledgement of the core MI processes along with the individual techniques which would be ideal for more experienced trainees.

Limitations and Future Directions

Several limitations were identified from the present research. The most pressing limitation was limited power as a result of small sample size. The number of participants at all stages of the training programme was significantly less than ideal, therefore power was reduced in all statistical tests. In two cases, potential analyses were omitted because the sample size was too small and power too low. Additionally, several participants did not complete the measures at both pre- and post-training which further reduced the sample size and therefore power.

It is acknowledged, however, that because the training programme was implemented only for CYFs staff, natural limits were placed on the size of the sample. In addition it was likely that some employees would have had to plan their workload around the training and therefore may have arrived late or had to leave training early.

A further limitation was the inability to pair pre- and post-training scores of the MIKAT. Due to confidentiality the scores could not be matched to particular participants, therefore affecting the analyses required to assess sensitivity to training. If scores could have been paired, simpler analyses that require smaller sample sizes to achieve adequate power would have been possible.

The final limitation was the absence of control groups included in the training programme. The single group pre–post design could not control for the possibility that gains made on both the MIKAT and the VASE-R at post-training were a result of having completed the measures at pre-training (i.e., a practice effect). Unfortunately, all other research utilising the MIKAT (Doran et al., 2011b; Hohman et al., 2009; Leffingwell, 2006) and the VASE-R (Bell & Cole, 2008, Doran et al., 2011a; Rosengren et al., 2008) had implemented similar designs. Therefore, there are no available research findings that can definitively rule out a practice effect with these measures.

It is recommended that the analyses included in the present research be replicated with the MIKAT and VASE-R, after the measures have been subjected to revision. However, an alternate study design and larger sample sizes should be used to account for the previously acknowledged limitations. A randomised controlled trial would be one ideal research design to use because it would ensure the inclusion of control groups. With regards to the sample sizes, in the present research, $N=54$ was required for 95% power in the paired-samples t-test. With the addition of control groups, an independent t-test would be required and, for 95% power, $N=88$ would be needed in each group. It is also recommended that the desired sample size be increased slightly to account for possible attrition. Therefore, future research should have a total participant pool of approximately $N=200$. It is acknowledged that recruiting and training that many participants would be a difficult task, particularly if the control groups would later receive training. However, employing similar methods to the present research—where training is administered to several smaller groups over an extended period of time—would make the task more feasible. It is also unlikely that the number of participants required could be found in one organisation; therefore, it may be necessary to recruit from several similar organisations.

Another potentially effective design could be cross-sectional: participants could be divided and compared depending on job title or experience level. This would eliminate the need for control groups and the requirement to carry out training at two different time points to ensure all participants are eventually trained. Again, sample sizes would need to be quite large depending on how many groups that participants are divided into. Even if they are simply divided into two groups of job titles or levels of experience, a total sample size similar to the randomised controlled trial (approximately $N=200$) would be required.

In summary, the conclusions made on the basis of this research suggest that the MIKAT and VASE-R are sensitive to the effects of training, and, therefore, have the potential to be assets to MI training and implementation plans. However, the results of the present research suggest that both the MIKAT and VASE-R may require refinement and further psychometric evaluation to improve their validity as measures of MI knowledge (MIKAT) and skill (VASE-R). The current study recommends possible methods of improvement for the measures. However, additional research is required to identify whether these recommendations would actually be effective. Due to the importance of effective implementation of methods such as MI, it is strongly recommended that future research refines and improves the MIKAT and VASE-R. If this were to occur, the number of resources that can reliably ensure effective training would therefore be increased.

References

- Baer, J. S., Rosengren, D. B., Dunn, C. W., Wells, E. A., Ogle, R. L., & Hartzler, B. (2004). An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence*, 73(1), 99–106.
- Baer, J. S., Wells, E. A., Rosengren, D. B., Hartzler, B., Beadnell, B., & Dunn, C. (2009). Agency context and tailored training in technology transfer: A pilot evaluation of motivational interviewing training for community counsellors. *Journal of substance abuse treatment*, 37(2), 191–202.
- Barnett, E., Sussman, S., Smith, C., Rohrbach, L. A., & Spruijt-Metz, D. (2012). Motivational Interviewing for adolescent substance use: a review of the literature. *Addictive Behaviors*, 37(12), 1325–1334.
- Barrett, R. D., McLellan, T. L., & McKinlay, A. (2013). Self versus Family Ratings of the Frontal Systems Behaviour Scale and Measured Executive Functions: Adult Outcomes following Childhood Traumatic Brain Injury. *PloS one*, 8(10), e76916.
- Becker, K. D., & Stirman, S. W. (2011). The science of training in evidence-based treatments in the context of implementation programs: Current status and prospects for the future. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(4), 217–222.
- Bell, K., & Cole, B. A. (2008). Improving medical students' success in promoting health behavior change: a curriculum evaluation. *Journal of General Internal Medicine*, 23(9), 1503–1506.
- Blase, K., & Fixsen, D. (2013). Core Intervention Components: Identifying and Operationalizing What Makes Programs Work. ASPE Research Brief. *US Department of Health and Human Services*.

- Bohman, B., Forsberg, L., Ghaderi, A., & Rasmussen, F. (2013). An Evaluation of Training in Motivational Interviewing for Nurses in Child Health Services. *Behavioural and Cognitive Psychotherapy*, 41(03), 329–343.
- Britt, E., & Blampied, N. M. (2010). Motivational interviewing training: a pilot study of the effects on practitioner and patient behaviour. *Behavioural and Cognitive Psychotherapy*, 38(2), 239–244.
- Brug, J., Spikmans, F., Aartsen, C., Breedveld, B., Bes, R., & Fereira, I. (2007). Training dieticians in basic motivational interviewing skills results in changes in their counselling style and in lower saturated fat intakes in their patients. *Journal of Nutrition Education and Behavior*, 39(1), 8–12.
- Cyf.govt.nz. (2014). *Youth justice - child, youth and family*. [online] Retrieved from: <http://www.cyf.govt.nz/youth-justice/index.html> [Accessed: 07 May 2013]
- Childers, J. W., Bost, J. E., Kraemer, K. L., Cluss, P. A., Spagnoletti, C. L., Gonzaga, A. M. R., & Arnold, R. M. (2012). Giving residents tools to talk about behavior change: A motivational interviewing curriculum description and evaluation. *Patient Education and Counseling*, 89(2), 281–287.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290 .
- Cucciare, M., Ketroser, N., Berg-Smith, S. M., & Chardos, J. (2012). Teaching Motivational Interviewing to Primary Care Staff in the Veterans Health Administration. *Journal of General Internal Medicine*, 27(8), 953–961.
- Daeppen, J. B., Fortini, C., Bertholet, N., Bonvin, R., Berney, A., Michaud, P. A., ... & Gaume, J. (2012). Training medical students to conduct motivational interviewing: a randomized controlled trial. *Patient Education and Counseling*, 87(3), 313–318.

- de Jonge, J. M., Schippers, G. M., & Schaap, C. P. D. R. (2005). The motivational interviewing skill code: Reliability and a critical appraisal. *Behavioural and Cognitive Psychotherapy*, 33(3), 285–298.
- Doran, N., Hohman, M., & Koutsenok, I. (2011a) Linking Basic and Advanced Motivational Interviewing Training Outcomes for Juvenile Correctional Staff in California. *Journal of Psychoactive Drugs*, 43(S1), 19–26.
- Doran, N., Hohman, M., & Koutsenok, I. (2011b). Motivational interviewing training in juvenile corrections: A comparison of outside experts and internal trainers. *Legal and Criminological Psychology*, 18, 262–273.
- El-Mallakh, P., Chlebowy, D. O., Wall, M. P., Myers, J. A., & Cloud, R. N. (2012). Promoting nurse interventionist fidelity to motivational interviewing in a diabetes self-care intervention. *Research in nursing & health*, 35(3), 289–300.
- Evangelini, M., Engelbrecht, S. K., Swartz, L., Turner, K., Forsberg, L., & Soka, N. (2009). An evaluation of a brief motivational interviewing training course for HIV/AIDS counsellors in Western Cape Province, South Africa. *AIDS care*, 21(2), 189–196.
- Evangelini, M., Longley, M., & Swartz, L. (2011). No reductions and some improvements in South African lay HIV/AIDS counsellors' motivational interviewing competence one year after brief training. *AIDS care*, 23(3), 269–273.
- Fichman, M., & Cummings, J. N. (2003). Multiple Imputation for missing data: Making the most of what you know. *Organizational Research Methods*, 6(3), 282–308.
- Forsberg, L. [Lars], Ernst, D., & Farbring, C. Å. (2011). Learning motivational interviewing in a real-life setting: A randomised controlled trial in the Swedish Prison Service. *Criminal Behaviour and Mental Health*, 21(3), 177–188.

- Forsberg, L. [Lisa], Forsberg, L. G., Lindqvist, H., & Helgason, A. R. (2010). Clinician acquisition and retention of Motivational Interviewing skills: a two-and-a-half-year exploratory study. *Substance Abuse Treatment, Prevention, and Policy* 2010, 5:8
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 857–871.
- Haeseler, F., AH, F., Pfeiffer, C., Walters, C., & Martino, S. (2011). Assessment of a motivational interviewing curriculum for year 3 medical students using a standardized patient case. *Patient Education and Counseling*, 84(1), 27–30.
- Hartzler, B., Baer, J. S., Dunn, C., Rosengren, D. B., & Wells, E. (2007). What is seen through the looking glass: The impact of training on practitioner self-rating of motivational interviewing skills. *Behavioural and Cognitive Psychotherapy*, 35(4), 431–445.
- Hohman, M., Doran, N., & Koutsenok, I. (2009). Motivational interviewing training for juvenile correctional staff in California: One year initial outcomes. *Journal of Offender Rehabilitation*, 48(7), 635–648.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The Sources of Four Commonly Reported Cutoff Criteria What Did They Really Say? *Organizational Research Methods*, 9(2), 202–220.
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment, Research & Evaluation*, 12(2), 1–11.
- Leffingwell, T. R. (2006). Motivational Interviewing Knowledge and Attitudes Test (MIKAT) for evaluation of training outcomes. *MINUET*, 13, 10–11.
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202.

- Lundahl, B. W., Kunz, C., Brownell, C., Tollefson, D., & Burke, B. L. (2010). A meta-analysis of motivational interviewing: twenty-five years of empirical studies. *Research on Social Work Practice, 20*(2), 137–160.
- Martino, S., Haeseler, F., Belitsky, R., Pantalon, M., & Fortin, A. H. (2007). Teaching brief motivational interviewing to year three medical students. *Medical education, 41*(2), 160–167.
- Miller, W. R. (1996). Motivational interviewing: Research, practice, and puzzles. *Addictive behaviors, 21*(6), 835–842.
- Miller, W. R., & Baca, L. M. (1983). Two-year follow-up of bibliotherapy and therapist-directed controlled drinking training for problem drinkers. *Behavior Therapy, 14*(3), 441–448.
- Miller, W. R., Hedrick, K. E., & Orlofsky, D. R. (1991). The Helpful Responses Questionnaire: A procedure for measuring therapeutic empathy. *Journal of Clinical Psychology, 47*, 444–448.
- Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. (2008). Manual for the motivational interviewing skill code (MISC): Version 2.1. *University of New Mexico, NM*.
- Miller, W., & Rollnick, S. (1991). *Motivational Interviewing: preparing people to change addictive behaviour*. New York: Guilford Press.
- Miller, W. R., & Rollnick, S. (2002). *Motivational Interviewing: Preparing People for Change*. New York; London : Guilford Press.
- Miller, W. R., & Rollnick, S. (2009). Ten things that motivational interviewing is not. *Behavioural and Cognitive Psychotherapy, 37*(2), 129–140.
- Miller, W. R., & Rollnick, S. (2012). *Motivational Interviewing: Helping people change*. New York; London: Guilford Press.

- Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *The American Psychologist*, 64(6), 527–537.
- Miller, W. R., Taylor, C. A., & West, J. C. (1980). Focused versus broad-spectrum behavior therapy for problem drinkers. *Journal of Consulting and Clinical Psychology*, 48(5), 590–601.
- Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of consulting and Clinical Psychology*, 72(6), 1050–1062.
- Mitcheson, L., Bhavsar, K., & McCambridge, J. (2009). Randomized trial of training and supervision in motivational interviewing with adolescent drug treatment practitioners. *Journal of Substance Abuse Treatment*, 37(1), 73–78.
- Moyers, T., Manual, J., Wilson, P., Hendrickson, S., Talcott, W., & Durand, P. (2008). A randomized trial investigating training in motivational interviewing for behavioral health providers. *Behavioural and Cognitive Psychotherapy*, 36, 149–162.
- Moyers, T., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skills code. *Behavioural and Cognitive Psychotherapy*, 31(2), 177–184.
- Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*, 28(1), 19–26.
- Moyers, T. B., Martin, T., Manuel, J. K., Miller, W. R., & Ernst, D. (2010). Revised global scales: Motivational interviewing treatment integrity 3.1. 1 (MITI 3.1. 1). *Unpublished manuscript, University of New Mexico, Albuquerque, NM.*

- Norman, G. R., & Streiner, D. L., (1998). Principal Components and Factor Analysis: Fooling Around with Factors. In G. R. Norman & D. L. Streiner, *Biostatistics, the Bare Essentials* (129–143). Hamilton, Ontario: B.C. Decker.
- Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, 56(1), 63–75.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32(3), 396–402.
- Pierson, H. M., Hayes, S. C., Gifford, E. V., Roget, N., Padilla, M., Bissett, R., ... & Fisher, G. (2007). An examination of the motivational interviewing treatment integrity code. *Journal of Substance Abuse Treatment*, 32(1), 11–17.
- Robbins, L. B., Pfeiffer, K. A., Maier, K. S., LaDrig, S. M., & Berg-Smith, S. M. (2012). Treatment fidelity of motivational interviewing delivered by a school nurse to increase girls' physical activity. *The Journal of School Nursing*, 28(1), 70–78.
- Rosengren, D. B., Baer, J. S., Hartzler, B., Dunn, C. W., & Wells, E. A., Ogle, R. (2005, revised 2009). Video assessment of simulated encounters (VASE-R): Administration and scoring manual. *University of Washington, Seattle, WA*.
- Rosengren, D. B., Hartzler, B., Baer, J. S., Wells, E. A., & Dunn, C. W. (2008). The video assessment of simulated encounters-revised (VASE-R): Reliability and validity of a revised measure of motivational interviewing skills. *Drug and Alcohol Dependence*, 97(1), 130–138.
- Schoener, E. P., Madeja, C. L., Henderson, M. J., Ondersma, S. J., & Janisse, J. J. (2006). Effects of motivational interviewing training on mental health therapist behavior. *Drug and Alcohol Dependence*, 82(3), 269–275.

- Smith, J. L., Amrhein, P. C., Brooks, A. C., Carpenter, K. M., Levin, D., Schreiber, E. A., ... & Nunes, E. V. (2007). Providing live supervision via teleconferencing improves acquisition of motivational interviewing skills after workshop attendance. *The American Journal of Drug and Alcohol Abuse*, 33(1), 163–168.
- Smith, J. L., Carpenter, K. M., Amrhein, P. C., Brooks, A. C., Levin, D., Schreiber, E. A., ... & Nunes, E. V. (2012). Training substance abuse clinicians in motivational interviewing using live supervision via teleconferencing. *Journal of Consulting and Clinical Psychology*, 80(3), 450–464.
- Stein, L. A., Colby, S. M., Barnett, N. P., Monti, P. M., Golembeske, C., & Lebeau-Craven, R. (2006). Effects of motivational interviewing for incarcerated adolescents on driving under the influence after release. *The American Journal on Addictions*, 15(1), 50–57.
- Stein, L., Lebeau, R., Colby, S. M., Barnett, N. P., Golembeske, C., & Monti, P. M. (2011). Motivational interviewing for incarcerated adolescents: effects of depressive symptoms on reducing alcohol and marijuana use after release. *Journal of Studies on Alcohol and Drugs*, 72(3), 497–506.
- Stein, L., Monti, P. M., Colby, S. M., Barnett, N. P., Golembeske, C., Lebeau-Craven, R., & Miranda, R. (2006). Enhancing substance abuse treatment engagement in incarcerated adolescents. *Psychological Services*, 3(1), 25–34.
- Tappin, D. M., McKay, C., McIntyre, D., Gilmour, W. H., Cowan, S., Crawford, F., ... & Lumsden, M. A. (2000). A practical instrument to document the process of motivational interviewing. *Behavioural and Cognitive Psychotherapy*, 28(01), 17–32.
- van Eijk-Hustings, Y. J., Daemen, L., Schaper, N. C., & Vrijhoef, H. J. (2011). Implementation of Motivational Interviewing in a diabetes care management initiative in The Netherlands. *Patient Education and Counseling*, 84(1), 10–15.

- Wahab, S., Menon, U., & Szalacha, L. (2008). Motivational interviewing and colorectal cancer screening: a peek from the inside out. *Patient Education and Counseling*, 72(2), 210–217.
- Walters, S. T., Alexander, M., & Vader, A. M. (2008). Officer Responses Questionnaire: A Procedure for Measuring Reflective Listening in Probation and Parole Settings. *The Fed. Probation*, 72, 67–70.
- Walters, S. T., Vader, A. M., Nguyen, N., Harris, T. R., & Eells, J. (2010). Motivational interviewing as a supervision strategy in probation: A randomized effectiveness trial. *Journal of Offender Rehabilitation*, 49(5), 309–323.
- Wolfe, H., Haller, D. L., Benoit, E., Bolger, K. W., Cancienne, J. C., Ingersoll, K. S., & Sharp, V. (2012). Developing PeerLink to engage out-of-care HIV+ substance users: Training peers to deliver a peer-led motivational intervention with fidelity. *AIDS care*, (ahead-of-print), 1–7.
- Young, T. L., & Hagedorn, W. B. (2012). The effect of a brief training in motivational interviewing on trainee skill development. *Counselor Education and Supervision*, 51(2), 82–97.

Appendix A

The MITI Coding Sheet¹

Motivational Interviewing Treatment Integrity Code (MITI)
Coding Sheet Revised June, 2007

Tape # _____ Coder: _____ Date: _____

Global Ratings

Evocation		1 Low	2	3	4	5 High
Collaboration		1 Low	2	3	4	5 High
Autonomy/ Support		1 Low	2	3	4	5 High
Direction		1 Low	2	3	4	5 High
Empathy		1 Low	2	3	4	5 High

Behavior Counts

Giving Information			
MI Adherent	Asking permission, affirm, emphasize control, support.		
MI Non-adherent	Advise, confront, direct.		
Question (subclassify)	Closed Question		
	Open Question		
Reflect (subclassify)	Simple		
	Complex		
	TOTAL REFLECTIONS:		

First sentence: _____

Last sentence: _____

¹ From "Revised Global Scales: Motivational Interviewing Treatment Integrity 3.1.1 (MITI 3.1.1)" by T.B. Moyers, T. Martin, J.K. Manuel, W.R. Miller, & D. Ernst (2007). Reprinted with permission.

Appendix B

The MIKAT²

MOTIVATIONAL INTERVIEWING QUIZ

The following statements are either factually true or false or consistent with ("true") or inconsistent with ("false") a motivational interviewing approach. Indicate your response by circling the appropriate item to the right.

1. Substance users must accept their problem (for example: "I am an alcoholic/addict.") before they can get help.	True	False
2. Denial is a characteristic of the disease of addiction.	True	False
3. Therapists' expectancies for their client's abilities to change have no effect upon whether change occurs.	True	False
4. Research has failed to find support the existence of an "addictive personality."	True	False
5. Substance users need to "hit bottom" before they can change.	True	False
6. If clients are resistant to talk about changing substance use, direct confrontation and persuasion are required to help the person change.	True	False
7. Resistance to talking about substance use is the direct result of denial, a symptom of the disease of addiction.	True	False
8. Counselors should emphasize personal choice over clients' behaviors, including substance use.	True	False
9. Substance abusers are generally incapable of making sound decisions in their current state of addiction.	True	False
10. Resistance is best thought of as a product of the interpersonal context in which it is observed.	True	False
11. Addicts and alcoholics are not capable of exerting control over their substance use behavior.	True	False
12. Readiness to make change is the client's responsibility – no one can help them until they decide they are ready.	True	False
13. The best way to motivate substance users is to help them resolve their ambivalence about change.	True	False
14. External pressure and consequences is the only way to make substance abusers change.	True	False
15. Which of the following are principles of a Motivational Interviewing approach to dealing with substance use? (select all that apply):		
<input type="checkbox"/> Breakdown denial	<input type="checkbox"/> Develop discrepancies	<input type="checkbox"/> Confront resistance
<input type="checkbox"/> Express empathy	<input type="checkbox"/> Acceptance of label ("alcoholic/addict") is required	<input type="checkbox"/> Educate about risks
<input type="checkbox"/> Maximize external pressure	<input type="checkbox"/> Use subtle coercion	<input type="checkbox"/> Support self-efficacy
<input type="checkbox"/> Roll with resistance	<input type="checkbox"/> Give direct advice	<input type="checkbox"/> Give clear consequences
<input type="checkbox"/> Require abstinence as only acceptable goal	<input type="checkbox"/> Encourage submission to disease	<input type="checkbox"/> Avoid argumentation

²From "Motivational Interviewing Knowledge and Attitudes Test for Evaluating Training Outcomes" in *MINUET* by T. R. Leffingwell (2006) p. 10–11. Reprinted with permission.

Appendix C

VASE-R Items with Sample Answers³

Item Scoring

Below you will find the item number, narrative that immediately precedes the respondent prompt, the prompts the item, scoring criteria and sample responses for each criterion. The samples are not meant to be exclusive, but instead to represent a category of response. There is a brief rationale given for the scoring of the 0 and 1 point samples.

Response Sample Item A.

Item Content

Pat: *You know, people are making a bigger deal out of this than is really needed. Everyone I know smokes a little pot now and then.*

Prompt: *Write a response that indicates you are listening.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) that are confrontational or likely to engender resistance. Responses that are either unclear or illegible receive a 0.

Maybe they're concerned because it is a big deal. (Begins counterargument)
It sounds to me like you are in denial about your drug addiction. (Diagnosing)
That's how people usually talk when they don't want to deal with a problem. (Dismissive)
Just because other people may do something doesn't mean it's OK for you. (Counterargument)
Do you know anything about addiction? (Likely to engender resistance)

1 Responses (e.g., statements or questions) that are neutral or inaccurate reflections

How often are you using marijuana now? (Question)
Let's talk about your marijuana use. (Neutral)

How much do you use, compared to these people you know? (Question)
How big of a deal do YOU think it is? (Question)
Sounds like you are concerned about some people you know. (Inaccurate Reflection)

2 Responses that are accurate reflections (or a summary). The content of accurate reflections might include: lack of personal concern about current drug use; external pressure from other people to present for treatment, normative statements about level of use of his peers. Accurate reflections might also include *implied* content that expresses understanding of his current concern, worry or wonder about change.

Your marijuana use doesn't seem like a big deal.
You're not using marijuana any more than lots of people you know. (Reasonable inference)

³ From "Video Assessment of Simulated Encounters(VASE-R) Administration and Scoring Manual" by D. B. Rosengren, J. S. Baer, B. Hartzler, C. W. Dunn, E. A. Wells, and R. Ogle (2005) pp. 7–24. Reprinted with permission.

*People are overreacting about your drug use.
So you use marijuana now and then, and you don't feel concerned about it.
Some people are giving you a bad time about using marijuana.*

Response Sample Item B.

Item Content

Pat: *I mean, what is their problem anyway? They're making a big deal out of nothing.*
Prompt: *Write a response that you think would be most helpful in this situation.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) likely to engender resistance.
 Responses that are either unclear or illegible receive a 0.

You're smoking a lot of marijuana, aren't ya? (Confrontation)
I guess you're in the denial stage of the addictive process. (Likely to elicit counterargument)
It doesn't sound like you're being very honest with your self. (Counterargument likely)
Finding faults in other people isn't going to help you get the support you need (Counterargument)
If everyone else is wrong about this, then why are you here? (Argumentative)

1 Responses (e.g., statements, reflections or questions) that are neutral (neither increasing or reducing resistance)

A lot of people feel that way when they first start talking about their drug use. (Neutral reframe)
Have you ever been in treatment? (Neutral question)
What do you think you need? (Collaborative, yet primarily neutral)
What do you know about treatment? (Unlikely to engender resistance)
Yeah, lots of people use marijuana. (Focus on one aspect, mainly neutral)

2 Responses (e.g., statements, reflections or open questions) likely to decrease resistance. The content of reflections might include: appreciation for his reluctance about need for change; recognition of perceived unfair treatment from other people; personal responsibility for change; include positive and negative content (double-sided reflection) or implied content (e.g., reflections about willingness to discuss drug use).

It doesn't seem fair to you the way some people are treating you right now.
You don't think being here is necessary.
Some people seem to be taking their own problems out on you.
It should really be up to you to figure out if any changes need to be made in your drug use.
You feel other people are making 'a mountain out of a molehill'; but to your credit, you were willing to come in and discuss your drug use.

Response 1.

Item Content

Lisa: *I don't know when things started to change, I mean somewhere along the way my drinking just sort of shifted. I did the usual partying in high school. No big deal. But then I went from having a couple drinks after work on a Thursday or Friday, to having something with dinner, a glass of wine, to having something before dinner, as well as with it.*

Prompt: *Write a response that indicates you are listening.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) that are confrontational or likely to engender resistance. Responses that are either unclear or illegible receive a 0.

I'm surprised you'd say it's no big deal. (Begins counterargument)

It sounds to me like your alcoholism is in mid-stage. (Diagnosing)

That's how people usually talk before they realize they have a problem. (Dismissive)

Often times people are the last ones to know there is a problem. (Counterargument)

Do you know anything about alcoholism? (Likely to engender resistance)

1 Responses (e.g., statements or questions) that are neutral or inaccurate reflections

What caused your drinking to change? (Question)

Let's talk about your drinking. (Neutral)

So, you've started to drink alone? (Question)

What do you think is going on there? (Question)

2 Responses that are accurate reflections (or a summary). The content of accurate reflections might include: a change in drinking; an increase in drinking; drinking in high school w/o consequences; and uncertainty about when drinking changed. Accurate reflections might also include *implied* content that expresses concern, worry or wonder about the change.

You're not sure when things shifted.

Before you realized it, your drinking increased. It caught you by surprise. (Reasonable guess)

You've noticed a change in your drinking habits.

You've seen a change in your drinking – it seems to you that you are drinking more.

So, you started to have a drink before dinner.

Response 2.

Item Content

Lisa: *I mean I am here, talking with you, which – no offense – but it isn't high on my list of favorite things to do. It says something ain't right.*

Prompt: *Write a response that indicates you are listening.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) that are confrontational or likely to engender resistance. Responses that are either unclear or illegible receive a 0.

You didn't end up here without reason. (Argumentative)

How do you explain this DUI, if you don't have a problem? (Calls for counterargument)
Yet if you don't drink and drive, you don't get DUIs. (Off target, likely to engender resistance).
If it causes problems, it is a problem. (Labeling often leads to resistance)
And drinking is one of your favorite things to do. (Not reasonably implied. Inaccuracy may lead to resistance or at least support for the status quo)

1 Responses (e.g., statements or questions) that are neutral or inaccurate reflections

Maybe this DUI is really a blessing in disguise. (Neutral and a little off target)
So, what led you to come in? (Question)
Have you had any other DUIs in the past? (Question)
What are some of your favorite things to do? (Question)
You are really confused. (Inaccurate Reflection)

2 Responses that are accurate reflections (or a summary). The content of accurate reflections might include being unhappy about being at the interview and awareness that something is not right. Accurate reflections might also include *implied* content that expresses concern, worry or wonder about drinking habits. Reflections that integrate content from the prior client statement are also considered accurate.

You don't want to be here, but there are some things that make you wonder.
You came even though you had mixed feelings because it feels like something isn't right.
You really aren't thrilled about being here.
You're a little bit concerned about the way things are going. (Reasonable guess)

Response 3.

Item Content

Lisa: *But, this DUI stuff is bullshit. I mean I was right at .08 and the cop was just hassling me. I think he wanted a date or something. So, my being here is way overkill, you know what I mean?*

Prompt: *Write a response that you think would be most helpful in this situation.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) likely to engender resistance. Responses that are either unclear or illegible receive a 0.

Do you realize you're blaming it on the cop when it's really a consequence of your own behavior? (Leads to counterargument)
The DUI may be bullshit, but you're going to need to face that it got you into trouble. (Confrontation)
If you were .08, you were impaired, regardless of the cop. (Confrontation)
I happen to believe those cops save a lot of lives by doing their job. (Counterargument)
You say its overkill, what if you had killed somebody? (Confrontation)

1 Responses (e.g., statements, reflections or questions) that are neutral (neither increasing or reducing resistance)

That's interesting that you see it that way. (Neutral)
So, what happened with the DUI? (Neutral Question)
So, you think this whole thing is bullshit? (Question)

You don't feel the reading was accurate. (Inaccurate reflection)
So where do things stand now?

- 2 Responses (e.g., statements, reflections or open questions) likely to decrease resistance. The content of reflections might include appreciation for her feelings and positive and negative content (double-sided reflection). Implied content might include reflections of her anger or her frustration. A question that shifts focus to an area of less resistance would be acceptable.

You feel you were given a bad deal.
So even though this feels unfair, I'm wondering what you'd like to talk about today that would feel useful. (Shifting focus)
So you think this whole thing is bullshit. (Reflection, as opposed to a question)
So you think maybe the cop had unethical motives.
So there are some things you're a bit concerned about, but you're still unhappy with the cop.

Response 4.

Item Content

Lisa: *On the other hand, I know that drinking and partying has gotten a little out of hand. I don't think I'm an alcoholic or anything. I mean my father was an alcoholic so I know what that looks like. I'm not there. But, I don't want to get there.*

Prompt: *Write a summary that you might say to Lisa and which touches on the things that you think are most important.*

Item Scoring

- 0 Responses that are not summaries or summaries that engender resistance. Responses that are either unclear or illegible receive a 0.

I guess I'm surprised at your conclusions. You've been drinking heavier. You have a family history of alcoholism. Now you've got a DUI. Don't you think that there might be something to all this stuff? (Summary that builds an argument)
It seems to me that you don't think you have a problem yet. (Not a summary)
It was painful growing up with an alcoholic. You don't want to do that to yourself. (Change talk, but not a summary).

- 1 Responses that are summaries and may include client ambivalence or change talk. Ambivalence refers to the client feeling two ways about a behavior. Change talk includes concern about the status quo, benefits of change, intention to change, and hope that change is possible.

So you're not sure that you need to be here. You think the cop gave you a raw deal. Your drinking is less than your Dad's, so you don't think you're an alcoholic. Yet, you've got this DUI so here you are. (Summary w/o ambivalence or change talk)
You sound concerned about alcohol in your life due to your father's alcoholism. You don't want to wind up like your father. (Doesn't include ambivalence)
You're worried about your drinking. Alcohol hurt your Dad and you don't want to repeat that. (Doesn't include ambivalence)
You saw what happened to your father and didn't like it. You don't want that to happen to you. (Doesn't include ambivalence)

- 2 Responses that are summaries and must include both client ambivalence and change talk

So, it sounds like you feel mixed – not really wanting to be here, but at the same time you don't want to end up like your Dad. This really concerns you.
So you are upset with the cop and having to come here. At the same time you've noticed some changes that you're not crazy about.
Let me see if I've understood. You are unhappy with the cop, the DUI process and being here today. You feel like the DUI was unfair and perhaps driven by some unethical motives on the cop's part. At the same time, you have some concerns because you've noticed some changes in your drinking. You fear becoming like your Dad. It hasn't escalated to that point and you don't want it to.

Response 5.

Item Content

Prompt: Write what you would say to Lisa that might elicit from her statements that support making healthy changes (e.g., concerns and/or recognition of problems, intention and/or optimism about prospective changes).

Item Scoring

- 0 Response engenders defense or justification of the status quo through antagonism, minimization of client concerns, or reinforcement of current behavior. Responses that are either unclear or illegible receive a 0.

So you won't even give this treatment idea a chance, will you? (Antagonistic)
You seem overly focused on the hassle of coming here today instead of what's really important. (Minimizes client concerns)
A DUI doesn't necessarily mean you are an alcoholic or anything like that (Reinforces status quo)

- 1 Response that is unlikely to elicit defense of the status quo or discussion of change. Content may be neutral, off-target, off-task or not supported by the client statements.

You might be surprised how many people are raised by an alcoholic parent. (Off-target/Off-task)
How are you getting along with your partner? (Neutral, off-target)
Talking to me isn't your favorite thing. (Presented alone, this focuses on things other than change)

- 2 Response clearly would elicit change talk. Change talk includes concern about the status quo, benefits of change, intention to change and hope that change is possible. Here are examples of change talk identified in Lisa's previous comments:

- *somewhere along the way my drinking just sort of shifted;*
- *it says something ain't right;*
- *On the other hand, I know that the drinking and partying has gotten a little out of hand.*
- *I'm not there. And, I don't want to get there, either.*

Here are sample responses to elicit change talk.

You mentioned that your drinking has shifted. Tell me more about that... (Elicits concerns about status quo)
If your drinking and partying have gotten a little out of hand, how might your life improve if you quit or reduced your drinking? (Elicits benefits of change)

Sounds as though you're not sure if you want to change this area of your life. If you did decide to try, how might you go about that? (Elicits optimism and hypothetical plan for change).

Response 6.

Item Content

Prompt: *There are many different directions to explore with Lisa. Select the question or statement that you think would be most helpful to explore with Lisa now, if you wanted to increase her motivation to change; then indicate why you chose that one in the space below.*

- _____ (1) What was your drinking like on the evening of the DUI?
- _____ (2) You indicated that being here is not a high priority for you. What would you rather be doing?
- _____ (3) You said, "something ain't right" about your situation. Tell me more about that.
- _____ (4) Tell me about your drinking patterns. How often and much do you drink?
- _____ (5) So what does an alcoholic look like to you?

What reason(s) led you to choose this item?

Item Scoring

The **correct answer** is #3, but the item is scored on a 3-point scale.

0 Incorrect response and the reason given for the response implies an intention to diagnose the problem, , educate the client or confront the client with her problem. The response may engender resistance. Responses that are either unclear or illegible receive a 0.

Respondent endorses # 1 or 4 and says:

"I need to know more about her drinking to determine if she really has a problem ."
(Diagnosis)

"I need to show her she has a problem." (Educating/confronting)

Respondent endorses #2 and says:

"I'm trying to press on her priorities a little bit to see if she's really thought it through." (Confront)

Respondent endorses #5 and says:

"I want to be able to show her that alcoholics are not just people who are homeless".
(Educating)

"I want her to see that she could be an alcoholic." (Educating)

1 Incorrect response, but the reason given for the response implies intent to better understand the client, reduce resistance or increase readiness for change. The response is unlikely to engender resistance.

Respondent endorses # 1 or 4 and says:

"I am trying to understand her perspective on her drinking."
"I want to elicit change talk (or reduce resistance by shifting directions)".

Respondent endorses #2 and says:

"I need to understand what's important to this client and how alcohol fits into that."

Respondent endorses #5 and says:

"This seems to be an important issue in her deciding if she has something to change so I want to see what she's thinking."
"I'm trying to avoid the labeling trap."
"I'm trying to build rapport."
"I want to understand what she is afraid of."

2 Correct response (reason is not scored)

Response 7.

Item Content

Ulysses: *I'll be honest with you, Doc. Things ain't going that great. But what I am doing, you know, I smoke a little crack, hang out with the fellas and we be drinking wine or whatever else we can get our hands on. And trying to come up with a hustle. Sometimes we beat somebody out of something or break into a car. Whatever we got to do. It's just the way it is. When you are out here and you don't have a home – You do whatever you got to do to survive.*

Prompt: *Write a response that indicates you are listening.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) likely to engender resistance.
 Responses that are either unclear or illegible receive a 0.

Sounds like your crack use and drinking are pretty out of control. (Overstates)
Sounds like it would be pretty hard to get any money together with all the crack you're smoking. (Likely to make client defensive)
It may not seem so bad now, but what about five years down the road? (Likely to engender resistance)
You're not going to find new friends, unless you get rid of the old ones first. (Advice giving when its not been asked for)
Hustling is just going to get you back into the joint. (Warning)

1 Responses (e.g., statements or questions) that are neutral or inaccurate reflections

How much crack are you using? (Question)
You're trying to put your life together since your release (Inaccurate).
Tell me more about your alcohol use (Question).
So, what do you mean days are better than nights? (Question)
When you say hustlin', what do you mean? (Question)

2 Responses that are accurate reflections (or a summary). The content of accurate reflections might include references to: being homeless; some ongoing substance use; feeling that alcohol is not an issue; hustling to make money; and the difference between nights and days. Accurate reflections might also include *implied* content that expresses acceptance of the situation.

You've accepted that this is the way things are going to be.
Days are better than nights.
You're just doing what it takes to survive.
Sometimes you do a little crack and drink a little wine.
Sounds like you are working hard to survive a difficult situation.

Response 8.

Item Content

Ulysses: *You know, I don't know, I don't know what good this is going to do me being here. Good thing is I got a place to lay my head. You feed me three times. And that's good 'cuz I was not getting that where I was. I also know that unless I am ready to make some changes, ain't nothing going to change. I don't know. I've done this treatment thing before. That's what I know.*

Prompt: *Write a response that you think would be most helpful in this situation.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) likely to engender resistance. Responses that are either unclear or illegible receive a 0.

You're still smoking crack, aren't ya? (Confrontation)
I guess that maybe you haven't hit bottom yet. (Likely to elicit counterargument)
I hate to say it, but it doesn't look like you're being very honest with your self. (Counterargument likely)
Insulting us isn't going to help any. (Respondent takes client statement as personal affront)
If treatment is crap, then why are you here? (Argumentative)

1 Responses (e.g., statements, reflections or questions) that are neutral (neither increasing nor reducing resistance)

A lot of people feel that way when they first come in. (Minor reframe, primarily neutral)
Have you been in treatment before? (Neutral question)
What do you think should be different with treatment? (Collaborative in nature, but mostly neutral)
What do you know about treatment? (Unlikely to engender resistance)
Yeah, honesty is an important part of this, isn't it? (Focus on a minor aspect is mostly neutral)

2 Responses (e.g., statements, reflections or open questions) likely to decrease resistance. The content of reflections might include: appreciation for his position; personal responsibility for change; include positive and negative content (double-sided reflection) or implied content (e.g., reflections of his doubts or fears about treatment).

You are very aware that you are the only who can make any changes
You're not sure treatment would be helpful.
You've done this before and you know its going to be hard.
You've learned from experience that it's really up to you.
It sounds like three hots and a cot can be pretty helpful sometimes, especially if things are not working out.

Response 9.

Item Content

Ulysses: *I be trying to take care of myself. Being out here is a trip. You know I got to watch out not only for the police, but I have to watch out for everybody else. Sometimes you don't know who's-who. So, it's not a movie, it's not wonderland, it's real. I also know I ain't with sleeping on the ground.*

Prompt: *Write a response that indicates you are listening.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) likely to engender resistance. Responses that are either unclear or illegible receive a 0.

The life you're describing sounds pretty bad. You know treatment can help you with some of this. (Caring advice, but still likely to engender resistance)

Wouldn't you like to get the cops out of your face? (Calls for a "Yes, but..." response)

Sounds like doing your thing isn't working. (Gentle, but confrontation)

And none of this has anything to do with your substance use? (Sarcasm)

Believe it or not most of this stuff is a result of your substance use. (Advising)

1 Responses (e.g., statements or questions) that are neutral or inaccurate reflections

Do you have any outstanding warrants? (Neutral, slightly off target)

How's your health doing these days? (Neutral, slightly off target)

I bet you have to pay attention to who you're dealing with. (Neutral, supportive)

So, you think the cops have it in for you. (Neutral, slightly off target)

2 Responses that are accurate reflections (or a summary). The content of accurate reflections might include references to: keeping to himself; being aware of people and surroundings; issues with the police; life not being a theme park; and sleeping on the ground. Accurate reflections might also include *implied* content that expresses dissatisfaction about his current lifestyle.

You find it very hard to trust anyone.

There are parts of your life you're not happy with.

It's hard living here and you have to watch out.

You're not too happy about sleeping on the ground.

It's hard always being on alert.

Response 10.

Item Content

Ulysses: *Things used to be pretty good. I thought. I had a little house. I had a woman and she was fine, and we was getting along and I got a couple years clean. Then the party came. You know I started getting loaded. Now she's gone, the little house is gone, and here I am.*

Prompt: *Write a summary that you might say to Ulysses and which touches on the things that you think are most important.*

Item Scoring

- 0 Responses that are not summaries or that are likely to engender resistance. Responses that are either unclear or illegible receive a 0.

Despite sleeping on the ground, losing your job, and your woman leaving, you still can't see that your substance use is a problem. (Not a summary and challenging)
So, what's it going to take for you to admit there is a drug problem? (Direct challenge)
So your life is hard (Nice reflection, but not a summary).

- 1 Responses that are summaries and may include either client ambivalence or change talk. Ambivalence refers to the client feeling two ways about a behavior. Change talk includes: concern about the status quo, benefits of change, intention to change, and hope that change is possible.

You've had good times and you've had bad times. Right now your life ain't great. (Change talk, but no ambivalence).
You used to have things you wanted: a woman, a job and a house. Those things are gone now. (Change talk, but no ambivalence)
So, you're sleepin' on the streets – fresh out of jail – and doing a little of this and a little of that; whatever it takes to survive. You're smoking a little rock and drinking a little, but you're not convinced it's a problem. Still, here you are at treatment. (Ambivalence, but no change talk)

- 2 Responses that are summaries and must include both client ambivalence and change talk

You're unsure you need treatment, but you are also not so crazy about how things are going. You know for sure things were better at one time. Right now, you're struggling with using drugs and finding it hard to trust others, too. You haven't made up your mind yet, but you are sure that only you can make the changes.
It's hard work out there and you've had some losses. You sometimes feel like there isn't much you can do about it. Yet, you know what it was like when you weren't using and you know something about what it will take to make changes.

Response 11.

Item Content

Prompt: Write what you would say to Ulysses that might elicit from him statements that support making healthy changes (e.g., concerns and/or recognition of problems, intention and/or optimism about prospective changes).

Item Scoring

- 0 Response engenders resistance talk (i.e., maintenance of the status quo) through antagonism, minimization of client concerns, or reinforcement of current behavior. Responses that are either unclear or illegible receive a 0.

Being homeless is just an excuse. You don't want to give treatment a real chance. (Antagonistic)
Don't you think that quitting drugs and alcohol will help stabilize the other areas of your life? (Minimizes client concerns)
Maybe you're right. Now may not be the right time for you to give treatment a try. (Reinforces status quo)

- 1 Response that is unlikely to elicit resistance talk or discussion of change. Content may be neutral, off-target, off-task or not supported by the client statements.

Life on the streets sound pretty rough. Had you stayed in any of the local shelters before you came in for treatment?(Off-target/Off-task)

Sounds like you are eating and sleeping better since you've been here. (Presented alone, this focuses on things other than change)

What would be a useful way to spend the rest of our time today? (Neutral statement, nondirective)

- 2 Response that would clearly elicit change talk. Change talk includes concern about the status quo, benefits of change, intention to change and hope that change is possible. Here are examples of change talk identified in Ulysses' previous comments:

- *Things aren't going that great;*
- *You know I got to watch out for not only the police, but I have to watch out for everybody else. Sometimes you don't know who's who;*
- *So, its not a movie, it's not a wonderland, it's real;*
- *Being out here is a trip;*
- *I ain't with sleeping' on the ground;*
- *Things used to be pretty good.*
- *The party came. You know I started getting loaded. Now she's gone, the little house is gone, and here I am.*

Like you said, it's no movie out there. You feel like you've had to be on the lookout for lots of things. (Elicits concerns about status quo)

You mentioned you've been eating and sleeping better since you came here. What are some other ways that remaining in treatment could be helpful for you? (Elicits benefits of change)

What would it take to begin to regain some of the important things you've lost in recent years? (Elicits optimism and hypothetical plan for change).

Response 12.

Item Content

Prompt: *Choose the statement or question that you think might be most helpful to explore with Ulysses, if you wanted to increase his motivation to change; then indicate why you chose that one in the space below.*

- _____ (1) You mentioned that being homeless is bad for your health and puts you at-risk for legal difficulties. If so, why not give treatment a chance?
- _____ (2) Don't you think your housing situation might improve if you stopped your drug use?
- _____ (3) So, even though all these tough things happened because of your drug use, you're still not sure you need treatment?
- _____ (4) What were things like when you were clean?
- _____ (5) What sorts of experiences have you had with treatment?
- What reason(s) led you to choose this item?*

Item Scoring

The correct answer is # 4, but the item is scored on a 3-point.

- 0 Incorrect response and the reason given for the response implies an intention to diagnose the problem, educate the client or confront the client with his problem. The response may engender resistance. Responses that are either unclear or illegible receive a 0.

Respondent endorses #1 and says:

I am trying to convince him treatment is not as bad as other things in his life.

(Persuasion)

Respondent endorses # 2 or 3 and says:

Because I need to help him see that drugs are hurting him. (Confrontation)

I am trying to break through his denial. (Confrontation)

Respondent endorses # 5 and says:

I'm trying to show him he's making a mistake in viewing all treatment as the same.

(Education)

- 1 Incorrect response, but the reason given for the response implies intent to better understand the client, reduce resistance or increase readiness for change. The response is unlikely to engender resistance.

Respondent endorses #1 and says:

I want to elicit the reasons for not wanting to change before I elicit the reasons to change.

Respondent endorses # 2 or 3 and says:

I am trying to understand how he came to these conclusions or these views.

Respondent endorses #5 and says:

I want to find out what he dislikes about treatment.

- 2 Correct response (reason is not scored)

Response 13.

Item Content

Bailey: *So, what are you gonna do, fix me? It's my parents that need fixing, not me. The losers.*

Prompt: *Write a response that you think would be most helpful in this situation.*

Item Scoring

- 0 Responses (e.g., statements, reflections or questions) likely to engender resistance. Responses that are either unclear or illegible receive a 0.

You've got to start looking at your own behavior, not your parents (Confrontation)

But they're only trying to help you because they love you (Advising)

But can't you see they do really love you? (Leads to counterargument)

That's pretty harsh, don't you think? (Leads to counterargument)

- 1 Responses (e.g., statements, reflections or questions) that are neutral (neither increasing or reducing resistance)

What's wrong with your parents? (Neutral question)

No. I'm not going to try to fix you. (Diverts, but does not return responsibility to client)

So, your parents are losers. (Neutral reflection)
I know you're mad at your parents, but I find name-calling rarely helps.

- 2 Responses (e.g., statements, reflections or open questions) likely to decrease resistance. The content of reflections might include: appreciation for her feelings of frustration and anger; her view of her parents; and concerns that you might try to fix or change her. A statement of client responsibility for change or a question that shifts focus to an area of less resistance would also be acceptable.

It's their problem, not yours.
You don't want me trying to get inside your head.
It seems totally wrong that they shipped you off to me to get fixed.
You sound angry with your Mom and Dad.
Deciding to fix or change anything is really up to you, but it seems like we are getting way ahead of ourselves. I'd like to get to know you a little before we talk about any fixing or changing.

Response 14.

Item Content

Bailey: *It's like they never let up. Maybe I wouldn't smoke pot if they'd just get off my back for ½ a second.*
Prompt: *Write a response that you think would be most helpful in this situation.*

Item Scoring

- 0 Responses (e.g., statements, reflections or questions) likely to engender resistance.
 Responses that are either unclear or illegible receive a 0.

Don't you think you're overstating it a bit? (Counterargument)
Is it really that bad? (Requests client to defend why it is bad)
You can't keep blaming them for your mistakes forever, you know. (Confrontation)
Let's get honest here. You're the one smoking dope now, not your dad. (Confrontation)

- 1 Responses (e.g., statements, reflections or questions) that are neutral (neither increasing or reducing resistance)

Why do you think they never let up? (Neutral question)
What sorts of things do your parents do? (Neutral)

- 2 Responses (e.g., statements, reflections or open questions) likely to decrease resistance. The content of reflections might include: appreciation for her feelings of constant pressure; attempts to change her against her will; and the possibility of change if only given some room. A statement that reframes the experience would also be acceptable.

It's hard having them on your back.
They're nagging doesn't bring out the best in you.
It seems like there's never a moment when they're not on your case.
It seems to you that their pressure makes you want to smoke more.
It's almost like you have to use something...
So, you might consider making some changes in your smoking, if you could just get some room to do it your way.

Response 15.

Item Content

Bailey: *Here's my Dad. Mr. Helpful, telling me how life is. I'm like, "Whatever, Dad". I mean he's so full of shit. He smoked pot when he was a kid. I just wish they'd leave me alone.*

Prompt: *Write a response that you think would be most helpful in this situation.*

Item Scoring

0 Responses (e.g., statements, reflections or questions) likely to engender resistance. Responses that are either unclear or illegible receive a 0.

But he got his act together and quit. What about you? (Challenge)

But there must have been a good reason why he stopped. (Calls for client to defend a position)

This is your chance not to make the same mistakes your dad made. (Warning/advising)

So, you want your Dad to cut you some slack, but you're not willing to cut him any.

(Challenge)

You know even though he's not doing it well, that he's just trying to help you. (Calls for a "Yes, but..." response)

1 Responses (e.g., statements, reflections or questions) that are neutral (neither increasing or reducing resistance)

You have something in common with him. (Neutral content)

Dads are full of advice, aren't they? (Neutral and slightly off target)

My Dad was the same way when I was your age. (Attempt to come alongside client may actually increase distance)

Some parents see things differently when they mature. (Generally neutral defense of Dad)

2 Responses (e.g., statements, reflections or open questions) likely to decrease resistance. The content of reflections might include: her sense of her father's hypocrisy; her feelings of frustration and anger with her father; and her desire to be left alone. A statement of client responsibility for change or a question that shifts focus to an area of less resistance or reframes the interaction would also be acceptable.

Just because he's your dad doesn't make him an expert on your life.

It seems phony to you that's he's telling you to stop something that he himself did.

And this is the last way in the world you want your Dad to try to help you; it's really unpleasant to you.

You'd really like to be free to make your own decisions and maybe you're own mistakes.

Response 16.

Item Content

Bailey: *They never believe what I say anyway, so why should I tell them anything. It's like they don't care about the truth or about me. It's just, like – I don't know – like, "Oh no. Our daughter's a drug addict. Let's go see a rent a friend". They don't know about the things that are really going on and I can't tell them. They'd totally flip out.*

Prompt: Write a summary that you might say to Bailey and which touches on the things that you think are most important.

Item Scoring

0 Responses that are not summaries or that are likely to engender resistance. Responses that are either unclear or illegible receive a 0.

They don't trust you. (Reflection)

So your parents see a problem in your using marijuana and your falling grades, but you don't. And even though you've got stuff going on that would freak your parents even more, you don't see that as a problem either. In fact, you don't see much of any problems.

I'm wondering how bad its gonna get before you to see a problem. (Confrontation)

You seem really angry and when you are this angry I don't think that anyone can help you very much. You need to think some about where you are, why your parents are so upset, and see if things can be better. I'd be happy to help you do that. (Confrontation)

1 Responses that are summaries and that may include either client ambivalence or change talk. Ambivalence refers to the client feeling two ways about a behavior. Change talk includes: concern about the status quo, benefits of change, intention to change, and hope that change is possible.

They don't care about you, so they sent you here to be fixed by me and that pisses you off. They don't know you well and they certainly don't know what's going on. (Summary w/o ambivalence or change talk)

So your parents keep trying all this stuff – jumping on your back, sending you here – because they think you have a drug problem. And you don't see it that way. You just wish they'd leave you alone and let you work out your own life. (Some change talk, no ambivalence)

2 Responses that are summaries and must include both client ambivalence and change talk

Everyone is telling you that you have a drug problem and you're feeling misunderstood and not heard. You feel kind of two ways about that. Part of you wants to be left alone and part of you would like someone to understand. There are some things that trouble you, but weed isn't one of them.

They're in a panic over your smoking some pot, which you disagree is something to panic about. There are other things going on and they don't know you well enough to figure that out, and that may pain you a little that they don't. You're certainly not going to tell them about that stuff because they'd go off the deep end, but there may be a part of you that wants to be understood and not so alone.

Response 17.

Item Content

Prompt: Write what you would say to Bailey that might elicit from her statements that support making healthy changes (e.g., concerns and/or recognition of problems, intention and/or optimism about prospective changes).

Item Scoring

0 Response engenders resistance talk (i.e., maintenance of the status quo) through antagonism, minimization of client concerns, or reinforcement of current behavior. Responses that are either unclear or illegible receive a 0.

All the teenagers I see who are using pot seem to complain about their parents ,that's just the marijuana talking. (Antagonistic)

I bet if you quit using marijuana, the other problems in your life wouldn't seem so unmanageable. (Minimizes client concerns)

It will be difficult for us to make much headway if you are unwilling to look at the problems your parents are concerned about. (Reinforces status quo)

- 1 Response that is unlikely to elicit resistance talk or discussion of change. Content may be neutral, off-target, off-task or not supported by the client statements.

You feel like your parents don't know what's going on. (Presented alone, this does not elicit change talk)

Do you have any friends that understand what you're going through with your parents? (Off-target/Off-task)

How do you want us to spend our time in these sessions? (Neutral statement, nondirective)

- 2 Response would clearly elicit change talk. Change talk includes concern about the status quo, benefits of change, intention to change and hope that change is possible. There is a limited amount of change talk in Bailey's comments and little of it has to do with her marijuana smoking. Most people may receive a one on this item. Here are examples of change talk identified in Bailey's previous comments:

- *Maybe I wouldn't have to smoke pot if they'd just get off my back for ½ a second;*
- *I mean its like they don't care about the truth or about me;*
- *They don't know about the things that are really going on. I can't tell them. They'd flip out.*

You are pretty worried about some of the things going on in your life. (Elicits concerns about status quo)

It sounds like you're in a difficult place, and that you feel all alone in this. (Elicits concerns about status quo)

If your parents did get off your back for a while, what would that be like for you? (Elicits benefits of change)

Response 18.

Item Content

Prompt: Choose the question or statement that you think might be most helpful to explore with Bailey, if you wanted to increase her motivation to change; then indicate why you chose that one in the space below.

_____ (1) What relationship do you see between your drug use and your dropping grades?

_____ (2) Your parents don't believe you and that bothers you. What upsets you about that?

_____ (3) So, how often and what sorts of drugs are you using?

_____ (4) Your situation does sound hard and would probably be easier if you were off the weed for a while. What do you think?

_____ (5) You said your parents would worry if they knew what was really going on. Bailey, can't you see that they're already worried and that's why they brought you here?

What reason(s) led you to choose this item?

Item Scoring

The **correct answer** is #2, but the item is scored on a 3-point.

- 0 Incorrect response and the reason given for the response implies an intention to diagnose the problem, educate the client or confront the client with her problem. The response may engender resistance. Responses that are either unclear or illegible receive a 0.

Respondent endorses # 1 and says:

"I need to connect the dots for her. She's not seeing how drug use is screwing up her life." (Education)

Respondent endorses # 3 and says:

"I need to show her she has a problem." (Education/Confrontation)

Respondent endorses #4 and says:

"Her drug use is affecting her thinking. I don't think she can make a sound decision until she's drug free." (Confrontation)

- 1 Incorrect response, but the reason given for the response implies intent to better understand the client, reduce resistance or increase readiness for change. The response is unlikely to engender resistance.

Respondent endorses # 1 and says:

"I want to elicit some change talk."

Respondent endorses # 3 and says:

"To give her a chance to not be judged for telling the truth."

Respondent endorses #4 and says:

"To present her with some options for changing her situation."

- 2 Correct response (reason is not scored)

Appendix D

VASE-R Scoring Guidelines⁴

General Scoring Rules

- Items are scored on a 3-level point system. Items are not intended to elicit qualitatively ‘good MI’ or ‘bad MI,’ but rather specific responses that are consistent with specific MI skill areas.
- To receive a score of 2, a response should *clearly* meet the outlined criteria. If it is *vague, unclear* or requires the rater to *infer meaning* it should receive a lower score. If a response is so unclear that it cannot be scored, the item receives a 0.
- When the *emotional tone is unclear* or unknown, and the tone could affect the scoring, the response should be given the benefit of the doubt and the higher score is applied.
- Raters will often encounter items where a respondent provides *multiple responses*. For example, a respondent may offer two reflections separated by “or.” If both response are rated above 0,, *raters should credit the score of the highest-rated response* (i.e., if a “1” and a “2”, then score the item “2.”). However, *if any of the responses are rated a 0*, then the response is considered ‘*spoiled*’ (see below) and the item *receives a 0*.

Subscale Scoring Rules

These descriptions contain item numbers, general scoring rules and score range for the subscales.

Reflective Listening.

This scale includes Responses 1, 2, 7, & 9.

- 0- Statements, reflections, or questions that are confrontational or likely to engender resistance
- 1- Statements or questions that are neutral; inaccurate reflections
- 2- Responses that are accurate reflections (or summaries)

Responding to Resistance.

This scale includes Responses 3, 8, 13, 14 & 15.

- 0- Statements, reflections or questions likely to engender resistance
- 1- Statements, reflections or questions that are neutral (neither increasing or reducing resistance)
- 2- Statements, reflections or open questions that are likely to decrease resistance.

Summaries.

This scale includes Responses 4, 10 & 16.

- 0- Responses that are not summaries or summaries that engender resistance.
- 1- Summaries that include either client ambivalence or change talk.
- 2- Summaries that include both client ambivalence and change talk

Change Talk.

⁴ From “Video Assessment of Simulated Encounters(VASE-R) Administration and Scoring Manual” by D. B. Rosengren, J. S. Baer, B. Hartzler, C. W. Dunn, E. A. Wells, and R. Ogle (2005) pp. 4–7. Reprinted with permission.

This scale includes Responses 5, 11, & 17.

- 0- Response likely to elicit resistance talk (i.e., talk that maintains the status quo).
- 1- Response that is neutral, inaccurate, off-target, off-task or not supported by the client statements (but essentially harmless).
- 2- Response likely to elicit change talk (i.e., problem recognition, benefits of or intent to change, ability to change, commitment to change).

Developing Discrepancy.

This scale includes Responses 6, 12, & 18.

- 0- Incorrect response and the reason given implies an intention to diagnose the problem, educate the client or confront the client with her/his problem. The response may engender resistance.
- 1- Incorrect response, but the reason given implies intent to better understand the client, reduce resistance or increase readiness for change. The response is unlikely to engender resistance.
- 2- Correct response (reason is not scored)

The correct answers for the items are: #6 = 3, #12 = 4 and #18 = 2.

Additional Decision Rules:

Reflective Listening, Summaries and Questions

- On responses that call for a reflection, summaries are also acceptable.
- A question is not a reflection. On responses that call for a reflection and a question is all that is given, then the response should be scored a “0” or “1”. However, adding a question after a reflection does not automatically spoil the response. Apply the “spoil” rule described below.
- Summaries normally include at least two sentences. A one-sentence summation could be regarded as a summary if it includes two or more ideas. However, a sentence with one idea followed by a question does not constitute a summary
- Open questions cannot be answered with a yes, no or frequency count. On items requiring an open question for a 2-point response, a closed question response would be scored a “1” or a “0” based on its content. The lone exception is a ‘scaling question,’ such as the following: “On a scale of 1-10, how important to you is it to...” This question can receive a 2-point score, if the content is otherwise appropriate to the client context.

Engendering Resistance

- Based on MI, statements that confront, label, blame, shame, dismiss, argue, and criticize should all be considered as likely to result in client defensive responses, and hence rated as engendering resistance.,
- Not every statement that potentially creates defensiveness in clients can be listed herein. We have attempted to list several for each question in the item-by-item scoring that follows. When considering whether a respondent response engenders resistance, think about what the next response from the client would be. What is most likely? This stance does require the

rater to take a reasonable guess about what a client would say next. Experience with MI is necessary for the reliable rating of this dimension.

- “*Inaccurate comments*” are answers where the respondent simply missed the target or misconstrued what was said. This would typically receive a score of 1. However, statements that are grossly inaccurate are likely to engender resistance as they imply the respondent was not listening to what the client said.
- On change talk items (5, 11 and 17), inaccurate, off-target or off-task comments are scored a “1” since they are considered to elicit neither change talk nor defensiveness.
- Stage-mismatched questions can also engender resistance. However, this decision calls for considerable judgment on the part of rater. Thus, we suggest that, for the purpose of this instrument, raters should disregard stage-mismatched questions as a source of resistance.

Unclear Responses

- As responses are scored on the basis of their content, raters should not be marked down for sloppy handwriting. However, a written statement that is totally illegible or unclear should be scored a “0”.
- When a response is partially illegible or somewhat unclear, raters should decipher legible and/or clear components and score these.
- Respondents may at times use arrows and other devices to indicate changes to answers. If clearly done, then items are to be scored accordingly. If the intent is unclear, the item should be scored “0.”

Extraneous Material and Spoiled Responses

- A response that contains the requisite elements for a correct response, but *also* contains additional elements that are likely to engender resistance is considered “*spoiled*” and the entire response is scored “0”. For example, an accurate reflection followed by a hostile question would be considered spoiled. “Your parent’s really frustrate you. Why can’t you give them a break?” illustrates this point. The “spoil” rule applies regardless of the order of MI-consistent and resistance-engendering material within a response.
- If additional elements do not “spoil” the response, the answer is scored according to the criteria for a “1” or a “2” point response. For example, a reflection followed by an open-ended question might be scored “1” or a “2”: “Your parent’s really frustrate you. What other things frustrate you?” The higher score (2) is rated in this case.

Implied Meaning and Inferred Meaning

- “*Implied meaning*” refers to respondent behavior. It is a *respondent’s reasonable guess* at what a *client or actor* thinks or feels (based on statements within the vignettes). This is acceptable behavior if the guess is reasonable (i.e., most respondents would agree to it).
- “*Inferred meaning*” refers to a rater behavior. The rater is unsure of what the respondent is trying to say in a response and guesses at what the meaning is. The rater should be very careful about inferring meaning in respondent comments when they are unclear. Instead, try to score the response based on the information that is present.

- If there is an omitted word/phrase or an incorrect word is included and this element is unessential or changes the sentence in a minor way, then meaning can be inferred.
- If there is an omitted word, phrase or contraction or an incorrect word is included and this element is essential or changes the sentence in a major way then meaning *cannot* be inferred. The response should be scored either as unclear or as written. For example, the respondent forgot to include a contraction. This response should be scored as written, since adding the contraction would substantially change the meaning of the sentence. “You are happy with your parents” versus “You aren’t happy with your parents” are two very different sentences.
- As the VASE-R is a skills-based assessment tool, respondents write responses from 1st person perspective as literally “what he or she would say to the client.” *Responses provided from 3rd-person perspective* (i.e., “Explore the client’s ambivalence about drinking.”) *do not receive points*, since it is not possible to know how the respondent would operationalize this response from the information provided.

Appendix E

The MIKAT Modified for Use with CYFs Staff

MOTIVATIONAL INTERVIEWING QUIZ

The following statements are either factually true or false or consistent with ("true") or inconsistent with ("false") a motivational interviewing approach. Indicate your response by circling the appropriate item to the right.

- | | | |
|---|------|-------|
| 1. Clients must accept their problem (for example: "I am an alcoholic/addict.") before they can get help. | True | False |
| 2. Denial is a characteristic of the disease of addiction. | True | False |
| 3. Therapists' expectancies for their client's abilities to change have no effect upon whether change occurs. | True | False |
| 4. Research has failed to find support the existence of an "addictive personality." | True | False |
| 5. Substance users need to "hit bottom" before they can change. | True | False |
| 6. If clients are resistant to talk about changing their behaviour, direct confrontation and persuasion are required to help the person change. | True | False |
| 7. Resistance to talking about substance use is the direct result of denial, a symptom of the disease of addiction. | True | False |
| 8. Counselors should emphasize personal choice over clients' behaviours, including substance abuse | True | False |
| 9. Substance abusers are generally incapable of making sound decisions in their current state of addiction. | True | False |
| 10. Resistance is best thought of as a product of the interpersonal context in which it is observed. | True | False |
| 11. Addicts and alcoholics are not capable of exerting control over their substance use behavior. | True | False |
| 12. Readiness to make change is the client's responsibility – no one can help them until they decide they are ready. | True | False |
| 13. The best way to motivate clients is to help them resolve their ambivalence about change. | True | False |
| 14. External pressure and consequences is the only way to make clients change. | True | False |

15. Which of the following are principles of a Motivational Interviewing approach? (select all that apply):

- | | | |
|--|---|--|
| <input type="checkbox"/> Breakdown denial | <input type="checkbox"/> Develop discrepancies | <input type="checkbox"/> Confront resistance |
| <input type="checkbox"/> Express empathy | <input type="checkbox"/> Acceptance of label ("alcoholic/addict") is required | |
| <input type="checkbox"/> Educate about risks | <input type="checkbox"/> Maximize external pressure | <input type="checkbox"/> Use subtle coercion |
| <input type="checkbox"/> Support self-efficacy | <input type="checkbox"/> Roll with resistance | <input type="checkbox"/> Give direct advice |
| <input type="checkbox"/> Give clear consequences | <input type="checkbox"/> Require abstinence as only acceptable goal | |
| <input type="checkbox"/> Encourage submission to disease | | <input type="checkbox"/> Avoid argumentation |